Transcriptome and genome specializations of *Oxytropis* (Fabaceae) arctic species

Annie Archambault
Department of Plant Science
McGill University, Montréal

McGill University, Macdonald Campus
21111 Lakeshore road, Sainte-Anne-de-Bellevue
Québec, Canada H9X 3V9

August 2013

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Thesis abstract

Molecular adaptations of arctic plants are not well understood, yet they are the basis for plant survival in a cold growth season, low light and nutritiously poor environment. This thesis consists in three studies: 1) a comparison of plantlet transcriptomes of arctic and temperate *Oxytropis* species; 2) detection of codons under selective pressure in genes from these species; 3) a phylogenetic study of the *Oxytropis* genus that characterizes recurrence of arctic lineages evolution within the genus.

The first study used cDNA library construction, suppression subtractive hybridization, followed by EST (Expressed Sequence Tags) sequencing and annotation, and resulted in a list of 489 differentially expressed genes. Arctic plantlets preferentially express genes of the "response to stimulus" and "ribosome biogenesis" categories, whereas temperate plantlets express genes of the "photosynthesis", "ribosome biogenesis" and "translation and nucleosome assembly" categories.

In the second study, genes were isolated from genomic DNA, codons under selection were detected, and the evolutionary relationships of gene copies were established for the PR-10, the ripening-related proteins, and the KS-dehydrin families. *Oxytropis* dehydrins are of a novel type (K-like Y_4 K S) and evolved freely, except for a few codons under negative selection that cluster in the Y-segment. The PR-10 is the only set of analyzed genes where evolving novel protein variant was once advantageous.

In the third study, the nuclear ribosomal ITS sequences from 97 specimens of 30 *Oxytropis* species was analyzed by phylogenetic and network approaches. The nine arctic species evolved from different temperate ancestors, through six lineages.

Résumé de la thèse

Les adaptations moléculaires des plantes arctiques sont encore mal comprises, elles sont pourtant à la base de la survie des plantes dans ces milieux dominés par une saison de croissance froide, une lumière faible et un environnement nutritionnellement pauvre. Cette thèse consiste en trois études 1) une comparaison du transcriptome des plantules d'espèces d'*Oxytropis* arctiques et tempérées; 2) la détection de pressions de sélection dans les gènes ces espèces; 3) une analyse phylogénétique qui caractérise la récurrence de l'évolution de lignées arctiques au sein du genre *Oxytropis*.

La première étude utilise la construction de banques d'ADNc, la technique de « suppression subtractive hybridization », suivi de séquençage et annotation de ESTs (Expressed Sequence Tags) et a résulté en une liste de 489 gènes exprimés de façon différentielle. Les plantules arctiques expriment de façon préférentielle les gènes de « réponse aux stimulus », de « biogénèse des ribosomes » alors que les plantules tempérées expriment les gènes de « photosynthèse », de « biogénèse des ribosomes » et de « traduction et assemblage des nucléosomes ».

Dans la seconde étude, les gènes ont été isolés de l'ADN génomique, les codons sous pression de sélection ont été détectés, et les relations évolutives ont été établies pour les membres des familles géniques de PR-10, « ripening-related proteins », et déhydrines . Les déhydrines sont d'un type nouveau (K-like Y_4 K S) et ont évolués librement, sauf pour un groupe de codons sous pression de sélection négative dans le segment-Y. Les PR-10 sont le seul ensemble de gènes analysés où l'évolution de nouvelles variantes de protéines a été avantageuse.

Dans la troisième étude, la séquence de l'espaceur ribosomal nucléaire ITS a été analysée par des approches phylogénétiques et de réseaux pour 30 espèces d'*Oxytropis* représentées par un total de 97 spécimens. Les neuf espèces arctiques ont évoluées de différents ancêtres tempérées, par six lignées différentes.

Acknowledgement

I would like to thank the FQRNT, the Fondation Desjardins (Girardin-Vaillancourt scholarships) and the Eben Hopson foundation who, by subsidizing my firsts years of doctoral studies, made this project possible; as well as the Centre Sève and Northern Scientific Training (NSTP) grants who funded the field trip to Baffin Island. I would like to thank collaborators for sharing *Oxytropis* seeds: Dr. Susan Aiken, Dr. Laurie Consaul, Dr. Danielle Prévost, Dr. Carolyn Parker, Dr. Tracy Sterling and Plant Gene Resources of Canada. I want to thank my colleagues, who kindly shared their knowledge during our discussions. I want to thank Lynn Gillespie and Susan Aiken, research scientists at the Canadian Museum of Nature for introducing me to the Arctic flora. I am grateful to my supervisor Martina Strömvik for her general guidance through this degree.

I would like to thank my close family and friends, for their support and encouragement.

List of abbreviations

3’ UTR	3' untranslated region (downstream of the gene coding sequence)
5’ UTR	5' untranslated region (upstream of the gene coding sequence)
ABA	Abscissic acid (plant hormone)
ADR6	Auxin Down Regulated 6
AFLP	Amplified fragment length polymorphism
ANOVA	Analysis of variance
BAC	Bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
BLASTN	Search nucleotide databases using a nucleotide query with BLAST
BLASTX	Search protein databases using a translated nucleotide query with BLAST
bp	Base pairs
cDNA	Complementary DNA
cDNA-AFLP	cDNA-amplified fragment length polymorphism
CNVs	Copy number variations
CO_2	Carbon dioxide
COR/LEA	Cold responsive/late embryogenesis abundant
cRNA	Complimentary RNA
CSBP	Cytokinin Specific Binding Proteins
dbEST	Expressed Sequence Tags database
DDBJ	DNA Data Bank of Japan
DEG	Differentially expressed genes
d_N:d_S	Ratio of nonsynonymous substitutions (d_N) relative to synonymous substitutions (d_S)
DNA	Deoxyribonucleic acid

dNTP	Equimolar mix of the four deoxyribonucleotides: dATP, dCTP, dGTP and dTTP
ELIP	Early light inducible proteins
EMBL	European Molecular Biology Laboratory
EST	Expressed sequence tags
FEL	Fixed Effects Likelihood
FQRNT	Fonds québécois de la recherche sur la nature et les technologies
GARD	Genetic Algorithm Recombination Detection
gDNA	Genomic DNA
GLM	General linear model
GO	Gene ontology
H1	One of the histone proteins
H2A	One of the histone proteins
H2B	One of the histone proteins
H3	One of the histone proteins
H4	One of the histone proteins
HSP	Heat shock proteins
ITS	Internal transcribed spacer
KEGG	Kyoto Encyclopedia of Genes and Genomes
LEA	Late Embryogenesis Abundant proteins
LEA D-11	Late embryogenesis abundant D-11
LHCAIII	Light-harvesting chlorophyll-a/b protein of photosystem I, type III
$MgCl_2$	Magnesium chloride
MIPS	Munich Information Center for Protein Sequences
MLP	Major Latex Proteins
MLP28	Major latex protein 28
mRNA	Messenger RNA

PAUP*	Phylogenetic Analysis using Parsimony and other methods
q-RT-PCR	Quantitative Real Time PCR
PCR	Polymerase Chain Reaction
PDF1	Plant Defensin 1
PR-10	Pathogenesis-Related class 10 proteins
RAPD	Random Amplified Polymorphic DNA
REL	Random effects likelihood
RT-PCR	Reverse Transcription PCR
SLAC	Single-likelihood ancestor counting
SNP	Single Nucleotide Polymorphism
SSH	Suppressive subtraction hybridization
STP	Specific tissue protein
tRNA	Transfer RNA

0.1 Thesis format

This thesis is written in the form of three manuscripts. Chapter 3, entitled "PR-10, defensin and cold dehydrin genes are among those over expressed in *Oxytropis* (Fabaceae) species adapted to the Arctic" has been published in Functional and Integrative Genetics in April 2011 (DOI: 10.1007/s10142-011-0223-6). Chapter 4, entitled "The Y-segment of novel cold dehydrin genes is conserved and codons in the PR-10 genes are under positive selection in *Oxytropis* (Fabaceae) from contrasting climates" has been published to Molecular Genetics and Genomics in December 2011 (DOI: 10.1007/s00438-011-0664-6). Chapter 5, entitled "Evolutionary relationships in *Oxytropis* species, as estimated from the nuclear ribosomal internal transcribed spacer (ITS) sequences point to multiple expansions into the arctic" has been published in Botany in August 2012 (DOI: 10.1139/b2012-023). Publications details such as keywords and publisher licenses for permission to include in the thesis are given in the connecting text of each chapter.

0.2 Contributions of authors

For the three manuscripts, I (Annie Archambault) am the first author and my supervisor Dr. Martina Strömvik is the senior author. Annie Archambault and Martina Strömvik designed the initial research project. Annie Archambault organized fieldwork and obtained necessary permits for fieldwork to the Canadian Arctic, conducted all molecular laboratory work, performed analysis, and drafted the manuscripts. Guidance of Dr. Martina Strömvik was essential and especially critical during the initial project elaboration and for writing manuscripts. Both authors have participated in writing the final version of the manuscripts.

0.3 Licenses

Plant collection in Nunavut was approved by Nunavut Research Institute license no. 0101505N-M to Annie Archambault. Use of biohazardous material for laboratory work was approved by McGill University Environmental Health and Safety license granted to Martina Strömvik's laboratory for the general project "Soybean promoter research". The publications based on published Chapters are included entirely here, with permission from Springer under license number 2695381485945 (Chapter 3), and license number 2835461220405 (Chapter 4), and with permission from NRC Research Press under license number 3006530369165 (Chapter 5). All license agreements are provided by Copyright Clearance Center.

0.4 Originality and contribution to knowledge

Results presented in this thesis were generated to shed light on a long-standing question: what make arctic plants able to colonize this extremely rigorous climate. Primary results are in the form of expression data and DNA sequences, which are analysed to understand genes' selective constraints and species evolutionary relationships. Contributions are both factual and conceptual.

- Factual contributions
    - DNA sequences: Materially, before the first set of sequences resulting from this project were deposited in GenBank, in August 2010, there were only 172 nucleotide sequences and no EST for *Oxytropis* species. Sequence data generated here added to the database 16 nuclear ribosomal internal transcribed spacer (ITS); 21 short genomic coding sequences used for the real-time RT-PCR experiment; 1245 ESTs longer

than 200 bp differentially expressed; and 105 genomic DNA sequences for five genes and gene families. These novel sequence data will now be freely available for various types of sequence analyses, to any researcher worldwide.

- Gene transcription data: Results from the cDNA library subtraction that resulted in 489 genes being differentially transcribed between arctic and temperate species, and from the real-time RT-PCR experiments on four genes, are the firsts report on gene expression on a true arctic plant.

- Conceptual contributions
  - Main conclusions from the comparative transcriptomics presented here were often unexpected. For instance, sets of genes related to ribosome biogenesis and assembly were found differentially transcribed in the arctic and in the temperate species but were not usually thought to participate in potentially adaptive processes.
  - Another unexpected finding was an increased expression for several response to stimulus genes (PR-10, KS-dehydrins, defensins, ELIP, LEA) but a decreased expression of others (ADR6, ripening related, acid phosphatase VSP) in arctic transcriptomes.
  - Expected and confirmed differential expression profiles were also found, for photosynthesis related genes.

- Gene and species evolution
  - A new type of KS-dehydrins was discovered in *Oxytropis* from this genomic DNA sequence analyses. The Y-segment of this protein, which has a still unknown function, is highly constrained compared to the rest

of the sequence, suggesting novel starting hypotheses for these enigmatic proteins.

- The phylogenetic analysis of 30 *Oxytropis* species presented here is the first to address the evolutionary history of the genus as a whole, a situation that was problematic for the Arctic, which flora is composed of lineages from different continents. Here, the sequence analyses from the ITS region confirmed the pattern suggested by taxonomy that the arctic *Oxytropis* lineages were formed through multiple expansions of temperate ancestors into the arctic.
- A scenario is presented for arctic *Oxytropis* evolution, combining data from the literature and from the present work, where arctic lineages formed after 4.6 Ma, shortly before to shortly after the onset of the arctic climate, at 3 Ma.

- More generally, results presented here illustrate the feasibility and usefulness of medium scale sequencing strategies to gain new insights into the genome of an organism poorly known at the molecular level.

Chapter 1

1 Introduction

Molecular data, mainly genomic nucleotide sequences and gene expression data, revolutionized the general field of comparative biology and opened the possibility to connect genotypes to phenotypes at a new scale. There were, as of 2011 when the results from this thesis were published, twenty one angiosperm plant genomes completely or almost completely sequenced and easily searchable through online databases (Duvick et al. 2008; Hellsten et al. 2010); and hybridization arrays, a common tool for gene expression analyses, were commercially available for at least thirteen plant species (Affymetrix 2011). Many of these plant species are economically important crops, but the list also includes other plants such as *Arabidopsis thaliana* and *A. lyrata, Populus trichocarpa, Medicago truncatula, Mimulus guttatus, Aquilegia coerulea, Brachipodium distachyon,* to name a few.

Divergence and polymorphism at the molecular level, between taxa or lineages, have been usefully explored to propose candidate genes for ecologically relevant traits in wild plants. For instance, divergence in gene expression profiles proved to be useful for discovering candidate adaptations of *Thlaspi caerulescens* to metalliferous soils rich in zinc (Filatov et al. 2006; Hammond et al. 2006; van de Mortel et al. 2006; Plessl et al. 2010). In another example, integrated data on polymorphism in gene expression and in allele frequencies among populations served as guide to understand adaptations of *Helianthus annuus* populations to drought and salinity stress, and to weediness habitat (Kane and Rieseberg 2007; Kane and Rieseberg 2008; Lai et al. 2008). Additionally, synonymous and non-synonymous substitutions in protein coding gene homologs can provide valuable data on evolutive constraints (Nielsen 2005).

Arctic flora is particularly fragile and threatened by rapid global changes (Solomon et al. 2007), but despite continued interest from the research community for almost a century (e.g. Wager 1938)), plant long-term survival in this harsh environment is still intriguing, because the characteristic features of arctic plants are not unique, they are also presents in other floras as well, but at a lower frequency. Furthermore, arctic plant biology has not yet reaped benefits from the great technological possibilities now available for decrypting molecular mechanisms underlying phenotypic traits.

Molecular data so far gathered for arctic plants mainly addressed phylogeny (evolutionary relationships), phylogeography (distribution of lineages within a species) or cytogenetic (chromosomes number) questions. This thesis presents studies that made use of molecular biology technologies to improve understanding of the molecular basis of plant adaptations to the arctic. We have used arctic and temperate species of the *Oxytropis* genus for these studies.

1.1 Hypotheses and Objectives

1.1.1 General Hypothesis

Molecular specializations of plants to the arctic can be inferred by comparing transcriptome (genes expressed) between an arctic plant and a close temperate relative; and by detecting selection at codons in nucleotide sequences for a subset of genes with differential expression.

General Objective: To test this general objective, the differentially expressed genes are identified between two arctic and two temperate *Oxytropis* spp. (Fabaceae), harvested at the plantlet developmental stage. Codons under negative and positive selection are then detected from the gene sequence isolated from genomic DNA form these four *Oxytropis* species.

1.1.2 Hypothesis 1

(H1.1) Among *Oxytropis* species, genetic divergence is low, and coding sequences of homologous genes are therefore expected to be conserved among species.

(H1.2) Arctic species do not form a monophyletic group within the *Oxytropis* genus.

Objective 1: To test specific hypotheses H1.1 and H1.2, genetic divergence and evolutionary relationships are estimated in several *Oxytropis* species from the commonly used genomic DNA sequence internal transcribed spacers (ITS) using networks. The steps undertaken towards Objective 1 are described in Chapter 5.

1.1.3 Hypothesis 2

(H2.1) Arctic and temperate *Oxytropis* species express different sets of genes from each other in their respective natural climate conditions and

(H2.2) The biological and potential adaptive meaning of this differential gene expression can be estimated by comparing *Oxytropis* sequences and expression data with public data on gene sequences, expression and ontologies.

Objective 2: To test specific hypotheses H2.1 and H2.2, coding regions of differentially expressed genes are identified using suppressive subtraction cDNA library technique followed by EST sequencing and sequence annotation by similarity searches and Gene Ontologies (GO) term assignment. The steps undertaken towards Objective 2 are described in Chapter 3.

1.1.4 Hypothesis 3

(H3) Arctic and temperate *Oxytropis* species regulate the identified sets of genes differently from each other in different climate conditions.

Objective 3: To test specific hypothesis H3, gene expression (by q-RT-PCR) of selected genes found differentially expressed after Objective 2 are characterized, in four *Oxytropis* species, under the two (arctic and temperate) climatic conditions. The steps undertaken towards Objective 3 are described in Chapter 3.

1.1.5 Hypothesis 4

(H4) Positive and negative selective pressures are not uniform between genes and along gene regions for differentially expressed genes.

Objective 4: To test specific hypothesis H4, selected genes found differentially expressed after Objective 2 are isolated and sequenced, from the genome of four *Oxytropis* species (two arctic, two temperate). This estimates presence and location of codons under negative and positive selection using divergence based methods. Sequence data resulting from Objective 4 are used for designing primers for testing H3. The steps undertaken towards Objective 4 are described in Chapter 4.

The thesis first provides an overview of the literature in Chapter 2, followed by three chapters of research results (Chapter 3, 4 and 5) in manuscript format. Chapter 6 is a general discussion and the final Chapter 7 is a summary and conclusions, including a statement of the originality and contribution to knowledge of the thesis.

Chapter 2

2 Literature review

2.1 Arctic botany

During the second half of the twentieth century, research on arctic plants life cycle was motivated by their potential in revegetating natural resources exploitation sites (Bliss 1971). Since the end of the twentieth century, it is primarily the response of arctic flora to climate change that has been investigated (Arft et al. 1999). It was demonstrated that although the warming outcome in terms of height and reproductive output is positive in most arctic species (Henry and Molau 1997), it is negative in other species (Totland and Alatalo 2002). Warming decreases species biodiversity after only three to six years (Walker et al. 2006) and short heat stresses deteriorate plant performance (Marchand et al. 2006). The arctic flora therefore seems particularly threatened by climate changes. Phylogeography of arctic species gained in popularity since the 2000, and focuses on the geographical distribution of genetic lineages (reviewed in Abbott and Brochmann 2003; Abbott 2008; Brochmann and Brysting 2008; Provan and Bennett 2008) and inference on location of refuge for living organisms during the last glacial maxima.

2.1.1 Definitions for the Arctic region

Arctic definitions (McGuire et al. 2006) relevant for the Arctic flora relate to the climate, to presence of permafrost, or can be summarize by ecozone classifications. A climate may be considered Arctic if the mean temperature of the warmest (summer) month is below 10°C (Natural Resources Canada 1995b).

According to this definition, the Arctic would extend almost up to Chicoutimi in the Quebec Province, and would cover a much larger area than when defined by presence of permafrost or by ecozones. The limit of continuous (>90% coverage) permafrost is a good indicator of the arctic constraints on plant life. According to the permafrost definition (Natural Resources Canada 1995a), the Arctic would also cover a large area, and extend almost up to Northern Ontario. For the rest of this thesis, the Arctic is all the ecozones classified as Arctic (Fig 2.1): Arctic Cordillera, Northern Arctic and Southern Arctic. The terrestrial ecozones summarize well the flora, wildlife, climate, landforms; and it is the most relevant definition for plant life. According to this definition (Natural Resources Canada 1995d), the Arctic is more restricted in size, it coincides with tundra land cover (Natural Resources Canada 1995c), and is mainly confined to the Northwest Territories and Nunavut, but reaches the most Northern tip of Quebec province, with 51° latitude being its most southern North American boundary.

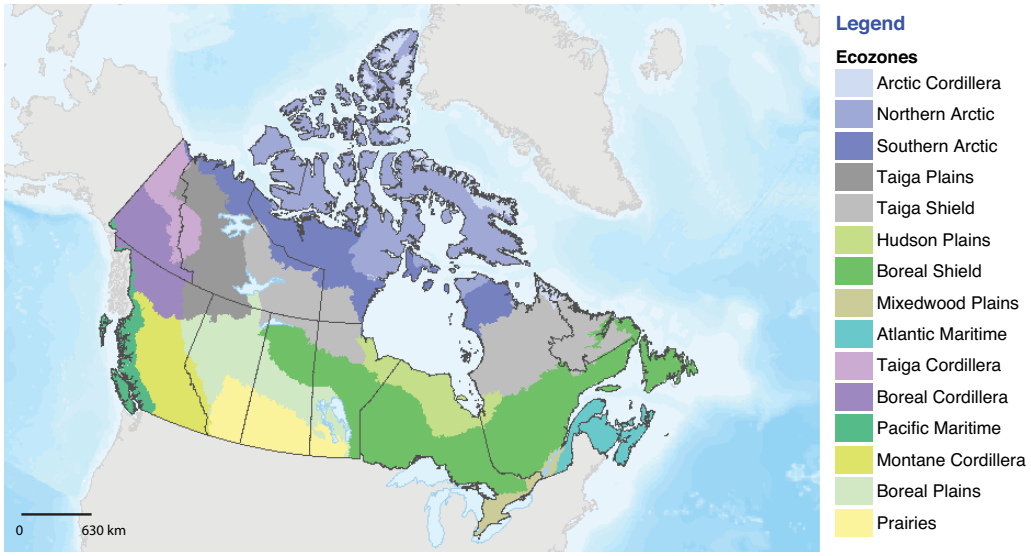


Figure 2.1 Terrestrial ecozones of Canada. In Canada, there are 15 terrestrial ecozones (Natural Resources Canada 1995d). The Arctic ecoregions are in lilac.

2.1.2 The constraints of arctic climate on plant establishment, and a note on plant strategies and adaptations

The rigorous climate of the Arctic is a harsh screen for plant life, as only 0.4% of the vascular plants species inhabit the Arctic (Svoboda 2009). The multiple constraints this climate imposes on plant growth and long-term survival are briefly reviewed in this section, and the characteristics and strategies that enable arctic plants to cope with these constraints are described.

The meaning of the term adaptation can be either strict or broad, depending on the context, and on the audience. The popular understanding refers to both the visible trait (adaptive trait) that appear to contribute to an organism' fitness, and to the process that leads to the appearance and maintenance of the trait (Collective 2012). Among the evolutionary biologists community, however, the definition of "adaptation" and "adaptive traits" is more strict. A trait can be qualified as adaptive when the ancestral and the derived trait, and the environmental changes influencing trait evolution are identified, and when traits performance can be measured (Lauder et al. 1993). Among population biologists, natural selection of advantageous alleles also comes into play to explain the process of adaptation (Barrett and Schluter 2008). Arctic plants' characteristics have not yet been scrutinized enough confirm that they are adaptive, and for that reason, the terms "adaptation to the Arctic" or "arctic adaptation" are avoided here; and the broader terms strategies or characteristics are preferred.

2.1.2.1 Short growing season

The short growing season, that lasts only from approximately July 20 to September 20 in the Low Arctic (9 weeks) and a few weeks less in the High Arctic (National Climate Data and Information Archive 2009) may be one of the most

severe constraints of the Arctic area on plant growth. This, combined to climatic severity may contribute to carbon shortage and induce general fragility of arctic plants (Wager 1938) and might explain that annuals life forms are almost absent in the Arctic (Bliss 1971; Billings 1987). To cope with very short growing season, arctic plants show several characteristics that are not unique, but occur in high frequency in the arctic flora. Prefloration or overwintering of floral buds, in which flower buds develop one to two years before flowering (Sørensen 1941; Billings and Mooney 1968; Bell and Bliss 1980) is one of these characteristics. At least twenty arctic species, including *Ranunculus sabinei*, *Papaver radicatum* and *Saxifraga cernua,* that are not true evergreens also keep their leaves for two seasons, a phenomenon termed wintergreen leaves (Bell and Bliss 1977).

Resuming growth extremely rapidly following snowmelt is another phenological strategy, that allows arctic plants to take advantage of all sunlight in the context of a short growing season (reviewed in Billings and Mooney 1968; Savile 1972; Billings 1987). Some phenological characteristics are phenotypically plastic. For instance, in the northern (but not arctic) annual biennial *Thellungiella*, long-term cold conditions promote earlier flowering when reproductive buds are still within the rosette, compared to growth in warmer conditions (Griffith et al. 2007)

Morphologically, the extensive root system developed by most arctic plants(Bliss 1971; Bliss and Gold 1999) is especially well suited for storing carbohydrates (Billings and Mooney 1968; Savile 1972) that enables this rapid spring growth. The short and cold growing season may explain many of the opportunistic characteristics observed in arctic plants, especially relative to reproduction. Arctic plants have no seed dormancy adaptation, and seeds may germinate readily, under light or darkness (Bliss 1971), when enough moisture and warmth (optimally up to 20°C) is provided (Billings and Mooney 1968; Bell and Bliss 1980; Bliss and Gold 1999). However, in terms of reproduction, although arctic plants set seeds by sexual

reproduction and insect pollination (Bliss 1971; Kevan 1972), seed set is often extremely low in the High Arctic and seedlings rarely survive the first winter (Bell and Bliss 1980; Bliss and Gold 1999). Therefore, vegetative and asexual reproduction is frequent in arctic plants (Bliss 1971).

The short growing season in the arctic, combined with cold temperature, allow only a very slow growth, when calculated on an annual basis. For instance, *Puccinellia vaginata* from the Canadian High Arctic flowers for the first time when approximately 26 years old (Grulke and Bliss 1988), and many species from this region appear to live at least 25 to 50 years (Bliss and Gold 1999). However, when compared on a daily basis, arctic plant productivity in the warmer and sunnier days of the growth season is comparable to that of temperate plants (Billings and Mooney 1968; Chapin 1983).

2.1.2.2 Light intensity

In the High Arctic (Fig 2. 1), illumination is continuous (24h) during the summer, but the light intensity is low, due to low solar angles with respect to the earth (Callaghan et al. 2005). Several arctic plants species and population adapted to this photoperiod, and require longer hours of light to flower (Billings and Mooney 1968; Bliss 1971; Teeri 1976). As a response to this constrain, arctic plants adopt growth forms that maximize interception of light, such as heliotropism (e.g. *Dryas integrifolia*, *Papaver radicatum*) where the plant turns to and with the sun (Kevan 1975).

2.1.2.3 Cold temperature

To human standards, the most striking characteristic of the Arctic is the extremely cold winter, where temperatures can reach an average of –25˚C to –35˚C (Aiken et al. 2007). Not surprisingly, arctic plants are freezing tolerant. For instance,

arctic *Silene acaulis* is freezing tolerant down to –30 C when acclimated, in mid winter (Junttila and Robberecht 1993). To plants standards, however, frequent summer frosts and cold temperatures during growing season, that remain below 10˚C in the Canadian Arctic Archipelago (Aiken et al. 2007), may impact survival even more profoundly. Accordingly, arctic plants such as *Silene acaulis* (Junttila and Robberecht 1993) and *Saxifraga cespitosa* (Robberecht and Junttila 1992) maintain efficient freezing tolerance at above zero growing temperatures. Arctic plants are also able to maintain active metabolism in a cold environment, well below 10 °C for the arctic populations of *Oxyria digyna* and at near freezing temperatures for some other species (Billings 1987). Arctic plants are notably plastic for this trait and modify their photosynthesis and respiration rates according to temperature (Billings and Mooney 1968; Pyankov and Vaskovskii 1994). The rate of incorporation of CO_2 into the products of photosynthesis is maximal at 15˚C for arctic *Caltha* and *Polygonum* species, and 25˚C for temperate species (Pyankov 1991). Indeed, plants of different climates exhibit similar respiration rate at average temperature of their respective habitat (Semikhatova et al. 2007).

Cold temperatures during growing season also affects plant indirectly, by limiting resource availability (Chapin 1983). An example of indirect effect of cold on arctic plant life is the permafrost that renders the soil anaerobic, decreasing soil microorganism biomass and activity, which in turn limits nutrient availability (Russell et al. 1940; Chapin 1983). Moreover, summer frosts, strong winds and low precipitation contribute to soil surface dryness and results in very high seedling mortality rate before the end of the fall (Bliss 1971; Bell and Bliss 1980; Bliss and Gold 1999).

Anatomical leaves features, such as thick leaves with large cells, characterize some arctic and alpine plants (Holzinger et al. 2007). The high alpine and arctic *Oxyria digyna* leaves also have chloroplast protrusions, which are specialized

thylakoid-free and chlorophyll-free structures (Buchner et al. 2007) that extend from regular thylakoid-containing chloroplast (Fig 2.2), and are in close contact with mitochondria (Holzinger et al. 2007). Arctic plants chloroplasts may also have few thylakoids per grana, low ratio of thylakoid to stroma volume (Miroslavov et al. 1996). The increase in organelles membranes and their proximity to mitochondria is suspected to enhance metabolic exchanges, and to participate in photorespiration as a protection against photoinhibition. Chloroplast protrusions are, however not generalized to all arctic species.

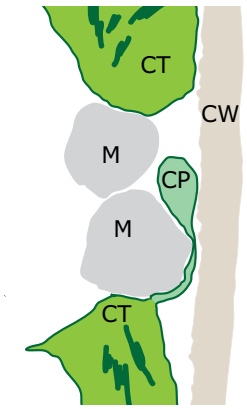


Figure 2.2 Drawing simplifying the cell ultrastructure of a *Oxyria digyna* leaf from Svalbard (Holzinger et al. 2007). The chloroplast protrusion (CP, thylakoid-free and chlorophyll-free (Buchner et al. 2007) is an extension of a thylakoid-containing chloroplast (CT), and is in close contact with mitochondria (M) in the vicinity of the cell wall (CW).

The complexity of the molecular responses to this abiotic stress is now better understood (Thomashow 2010) in crops. Among the many genes with suspected roles in cold acclimation and freezing tolerance, key genes are the CBF transcription regulator genes, which are induced by cold, and whose gene products (proteins) in turn, induce the expression of more than a thousand genes that have a positive

effect on cold tolerance (Thomashow 2010). The transcript profile of only a handful of these genes was surveyed in arctic plants. Within the Brassicaceae *Draba*, it was shown that temperate and arctic species exhibit a subtle difference, where temperate species show a faster induction of CBF genes than does the arctic ones (von Meijenfeldt 2010). This difference was, however, not reflected in the expression pattern of a COR15a gene (von Meijenfeldt 2010), which in *Arabidopsis* is quickly induced after the CBF genes are expressed.

2.1.2.4 Low diversity in the Arctic is a consequence of cumulative constraints

Species diversity is lower in the Arctic than at the temperate latitudes (Callaghan et al. 2004). This observation is a likely consequence of the recent colonization of the Arctic region that took place less than 14 000 years ago (Hulten 1968; Abbott and Brochmann 2003; Aiken et al. 2007), of the limited sexual reproduction, of low seedlings survival rate (Bell and Bliss 1980; Bliss and Gold 1999), in addition to the constraints discussed above. These constraints logically limit Arctic biodiversity, though Arctic plants show a surprising high level of biological diversity, even for populations with little morphological differences (Grundt et al. 2006). The high frequency of recent and high-level polyploids in the arctic flora (Brochmann et al. 2004; Brochmann and Brysting 2008), which increases allelic diversity within one individual genome by fixed heterozygosity, could be a key feature of arctic plants to avoid genetic and ecological depauperation (Brochmann et al. 2004).

Arctic plants therefore have lots to teach plant biologists, and prompted (Billings and Mooney 1968) to state that:

> "by studying arctic and alpine plants we can hope to learn how this relative handful of species in the world's flora has succeeded not only in surviving low temperatures during dormancy but in manufacturing relatively large amounts of food at low temperatures in very short periods of time."

2.1.3 *Oxytropis* sp. (Fabaceae) as a model system for discovering potentially relevant molecular specialization of arctic plants

In order to detect novel and potentially relevant features to explain long-term survival of arctic plants in this harsh climate, species of the *Oxytropis* genus were selected as a model for several reasons. There are several *Oxytropis* species with a true arctic distribution that can be compared to related *Oxytropis* species growing in temperate regions (Hulten 1968; Welsh 1991; Yurtsev 1999; Aiken et al. 2007). The seed production and seed germination of *Oxytropis* facilitate transportation and laboratory manipulations. Moreover, *Oxytropis* is taxonomically placed in the Papilionoideae subfamily of the legume family (Fabaceae) where numerous genomic data are available. Among available data are genomic DNA sequences from *Medicago truncatula* (Young et al. 2011) and *Glycine max* (soybean; (Schmutz et al. 2010), integrated in the Phytozome database web interface (Hellsten et al. 2010), as well as an impressive quantity of Expressed Sequence Tags (ESTs) for soybean (Shoemaker et al. 2002; Vodkin et al. 2004). The available molecular data on related species would facilitate the molecular experiments in *Oxytropis*.

Consumption of *Oxytropis* spp. and *Astragalus* spp. (known as locoweed) has been associated with locoism, a ruminant animals disease caused by the toxic alkaloid swainsonine (Stegelmeier et al. 1999). The fungal endophyte, *Undifilum oxytropis*, found in some *Oxytropis* and *Astragalus* species, is responsible for the synthesis of swainsonine (Braun et al. 2003; Valdez Barillas et al. 2007; Ralphs et al. 2008; Cook et al. 2009; Pryor et al. 2009). Distribution of the swainsonine-producing

fungus is not known in the arctic, but it is presumed not to be common since arctic muskox (Mulder and Harmsen 1995) graze *Oxytropis* species. Although *Oxytropis* species are not appealing for agriculture purposes in temperate areas because of their potential toxicity, they are valuable for revegetation, especially in boreal and arctic climates (Klebesadel 1993).

Figure 2.2 *Oxytropis maydelliana* growing by a telephone pole in August 2005 in Kimmirut (Nunavut). Photograph by Annie Archambault.

Oxytropis are insect cross-pollinated plants, they seem unable to self-pollinate autonomously (Kudo and Harder 2005), but this may not be accomplished by a genetic self-incompatibility system (Artyukova et al. 2004). Flowering in temperate *Oxytropis* lasts for 9 to 10 days in Kananaskis valley (Alberta, Canada) where *O. sericea* is mostly visited by *Bombus flavifrons*, and *O. splendens* by *B. californicus* and *B. nevadensis* (Kudo and Harder 2005). In the Arctic, *B. polaris* is a common pollinator (Kevan 1972) for *Oxytropis* species. *O. nigrescens* in Central Alaska onset of flowering can occur in early June, and seed dispersal may begin in late July (Murray and Miller 1982).

2.1.3.1 *Oxytropis* taxonomy and distribution

The *Oxytropis* genus is mainly distributed in the northern hemisphere, and is especially rich in species diversity in Eurasian mountains (Malyshev 2008) and in Beringia, which might have served as refugia in the last glacial era (Yurtsev 1999). Of the 64 (Yurtsev 1999) to 71 (Elven 2007) species and subspecies described with occurrence in the Arctic, only 8 occur in the Canadian Arctic (Aiken et al. 2007). This number of species is enough to place *Oxytropis* as the legume genus with highest species diversity in the Canadian Arctic (Aiken et al. 2007). No *Oxytropis* occur in High Arctic regions (Ellesmere, North Greenland, Svalbard).

The number of recognized taxa in *Oxytropis*; 300 (Langran et al. 2010) to 450 species (Malyshev 2008); should decrease markedly when the whole genus will be investigated experimentally (Elven 2007). The discrepancy in number of described species is partially explained by the many differences in treatments between the North American (Welsh 1991) and Asian Floras (Yurtsev 1999; Malyshev 2009). Russian authors typically place small taxa at the species rank, and North American authors group them within larger species, at the subspecies or variety level (Aiken et al. 2007; Elven 2007). Moreover, morphological intermediates, perhaps resulting from hybridization and allopolyploidization (genome duplication following a hybridization event), complicate establishment of subspecies to section boundaries (Welsh 1991; Yurtsev 1999). As a result, many species are rare and a clear morphological separation of closely related species sometimes lacks. Chromosome number, a useful characteristic for species definition, is also variable for some *Oxytropis* species (Ledingham 1957; Ledingham 1960; Ledingham and Rever 1963; Elven 2007). The basic chromosome number for *Oxytropis* is $x = 8$ ($2n = 2x = 16$);

and there are high polyploids with 96 chromosomes (Elven 2007). In general however, despite some difficulty in assigning taxonomic rank to closely related species, the circumscription of the 6 subgenera and 25 sections within *Oxytropis* seems clear and not disputed (Malyshev 2008; Langran et al. 2010).

2.1.3.2 Molecular diversity of *Oxytropis* species

Surveys of genetic differentiation, investigated in different parts of the world for rare *Oxytropis* species, paralleled the common lack of sharp morphological separation between closely related taxa. Patterns of low differentiation between sister taxa, high heterozygosity, high within population diversity and low among population diversity were also revealed (Jorgensen et al. 2003; Artyukova et al. 2004; Chung et al. 2004; Kholina et al. 2004; Schonswetter et al. 2004; Kholina et al. 2009). In some cases, subspecies were not genetically different from the larger specie complex (Jorgensen et al. 2003; Schonswetter et al. 2004), but genetic differences were noted in other instances (Chung et al. 2004).

The rare Switzerland subspecies of *O. campestris* subsp. *tirolensis* with restricted distribution was not different genetically from the widespread species when analyzed with AFLP (Amplified Fragment Length Polymorphism) markers (Schonswetter et al. 2004). Evolution of a genetically insignificantly differentiated phenotype during the recent migration is the most likely explanation to this unexpected result. A similar case occurs in Alaska, where genetic markers (RAPD, Random Amplified Polymorphic DNA) grouped individual of *O. arctica* and *O. campestris* many subspecies according to their geographic origins rather than their taxonomic recognition (Jorgensen et al. 2003). Populations of another rare species, *O. campestris* var. *chartaceae* from Wisconsin USA, analyzed by AFLP genetic

markers, showed relatively high heterozygosity, high within population diversity, but low among population diversity. Phylogenetic analyses performed on the AFLP data confirm this rare taxon status, separate from other *O. campestris* taxa (Chung et al. 2004). The rare autotetraploid *O. chankaensis* species, located at the border of China and far East Russia (Primorye), has high heterozygosity levels (measured with isozymes) for a rare endemic (Kholina et al. 2009), and the most northern and southern populations lacks diagnostic RAPD molecular markers (Artyukova et al. 2004), although they showed some genetic isolation. These described patterns of genetic differentiation are consistent with the *Oxytropis* outcrossing breeding system. Overall, a general picture of the evolutionary relationships within the *Oxytropis* genus, and especially between species from different continents, is still missing from the literature.

2.1.4 Diversity at the molecular level

Variations at the molecular level can be used to explain diversity among species and polymorphism among populations (among many reviews: Storz 2005; Whitehead and Crawford 2006b; Jensen et al. 2007; Karrenberg and Widmer 2008; Tenaillon and Tiffin 2008; Amtmann 2009; Gossmann et al. 2010; Siol et al. 2010). However, as for any other type of data in comparative biology, the challenge of identifying the adaptive differences among the ones carrying no selective advantage is enormous, which brought (Storz 2005) to say:

> "The identification of causative sequence variants represents a monumental challenge even in model organisms."

Several types of variation at the molecular level are available for comparison among population, lineages or species (Table 2.1): They can be in the form of

genomic DNA sequences (Anisimova and Liberles 2007), allelic or marker frequencies (Storz 2005; Siol et al. 2010), epigenetic modifications (Martienssen et al. 2005), or transcriptome characterization (Whitehead and Crawford 2006a). These data are mainly generated by various types of sequencing technologies, microarray hybridizations and fragment length analysis.

2.1.4.1 Transcriptome variations among taxa can be discovered by a variety of techniques

The transcriptome is the collection of all transcribed genes of a genome in an organism, organ, or cell at a given time, including messenger RNA (mRNA) of protein coding genes transcripts, transfer RNA (tRNAs), ribosomal RNA (rRNAs), and non-coding RNA. Difference in gene expression between species has long been suspected to be involved in species physiological, morphological and behavioral differences (King and Wilson 1975). One of the major challenge of the comparative transcriptomics approach is to identify which few genes could have evolved by natural selection among the often larger proportion of genes with differential expression that may be selectively neutral (Whitehead and Crawford 2006a).

Table 2.1 Overview of the various molecular data available for comparative studies among population or species, with application in plants.

Type of molecular variation	Brief description	Technology	Example in plant biology
[a] Sequence divergence	Variation in non-synonymous to synonymous substitutions ratio (d_N:d_S) in coding sequences among homologous genes.	Coding region sequencing (Sanger or short-read) from a set of paralogous or orthologous genes.	Discover one yet uncharacterized gene, expressed following biotic stress, and under positive selection among cereals genomes (Zamora et al. 2009). The recessive plant gene eIF4E, conferring plant virus resistance, has one codon under strong positive selection (Cavatorta et al. 2008). In species with different flower color, no evidence of positive selection is detected in genes of the anthocyanin genes pathway that control pigmentation (Toleno et al. 2010).
Population differences	Analysis of frequency distribution or patterns of nucleotide variation segregating at a locus between and within populations of a species.	Molecular markers, from fragment analysis (AFLP, microsatellites) or sequencing (Sanger or short read). Often associated with high-density genetic map. Genomes scans	There is almost no variation within a 500 kb region between 20 *Arabidopsis thaliana* ecotypes (Clark et al. 2007). One AFLP locus is under selection and its allele frequency is negatively correlated to temperature in the establishment year (Jump et al. 2006). Several SNPs correlate to geography, temperature, growing degree-days, precipitation and aridity, in a genome-wide comparison of SNPs among 54 populations of *Pinus taeda* L (Eckert et al. 2010).
[a] Gene expression	Comparison in mRNA levels in the whole plant or in one organ between	Hybridization on arrays (macroarrays, cDNA microarrays, oligonucleotides	Genome doubling has a small effect on gene expression changes compared to species hybridization (Wang et al. 2006). Out of 7500 genes, 60 show different mRNA levels among different accessions of

		arrays). Species-specific or cross species arrays.	*Arabidopsis thaliana* (Chen et al. 2005).
[a] Gene expression	different species or lineages.	cDNA libraries sequencing by suppression subtractive hybridization (SSH) or fragments of cDNA-AFLP comparison.	There is a consistent differential expression of four genes of the anthocyanin pathway between red and white grape berries, as detected by SSH (Ageorges et al. 2006). There are 18 genes differentially expressed carrying various functions between two *Phaseolus vulgaris* cultivars with contrasting drought tolerance, as detected by SSH (Montalvo-Hernandez et al. 2008).
Epigenetic modification	Variation in DNA cytosine methylation	gDNA treated with methylation-dependent restriction enzyme or bisulfite sequencing	Methylation in genes is highly polymorphic between two *Arabidopsis thaliana* ecotypes, but transposable elements are heavily methylated in both ecotypes (Vaughn et al. 2007).
	Histone modification (histone deacetylation or histone H3 Lys9 methylation $H3K9^{Me}$ repress gene expression)	Affinity purification or chromatin immunoprecipitation combined with hybridization on genomic arrays.	Unpublished results (Grossniklaus & Baumberger) reveal that in the *Mimulus aurantiacus* species complex, phenotypic transition from small red to large yellow flowers occurs during the life of individual plants and is controlled by epigenetic factors (Siomos 2009).
Variation in gene family size	The number of genes in a gene family differs among individuals of a same species, and between different species.	Complete genome sequencing (Hahn et al. 2005). It is termed Copy Number Variation in human genomics.	Tandemly arrayed genes are common in the *Arabidopsis* and in the rice genomes (Rizzon et al. 2006), and polyploidy is prevalent in plant genomes (Adams and Wendel 2005; Doyle et al. 2008), reviewed previously (Flagel and Wendel 2009; Freeling 2009).

[a] These two types of molecular variation are analyzed in the present thesis.

Most reports on variation in gene expression among plant population or related species were carried out on model species (Taji et al. 2004; Chen et al. 2005; Kliebenstein et al. 2006; Engelmann et al. 2008) and their close relatives. The most common technique for evaluating whole transcriptome variation among species or population was microarray analyses often using cross-species hybridizations, where the cRNA applied to the array is not from the same species as the oligo or cDNA spots on the array. Examples of taxa analyzed with the cross-species hybridizations technique are *Arabidopsis* relatives (Taji et al. 2004; Filatov et al. 2006; Hammond et al. 2006; van de Mortel et al. 2006; Wang et al. 2006; Broadley et al. 2008; Morinaga et al. 2008; Voelckel et al. 2008), *Helianthus* species or populations (Lai et al. 2006; Lai et al. 2008), *Senecio* species (Hegarty et al. 2009), *Picea* populations (Holliday et al. 2008), *Populus* population (Brosche et al. 2005; Fluch et al. 2008), *Medicago* species (Chen et al. 2008a), *Spartina* species (Chelaifa et al. 2010a; Chelaifa et al. 2010b) or *Zea mays* lines (Casati et al. 2006). Microarray analyses is a powerful technique, with repository databases (Barrett et al. 2007) and established statistical analysis that can estimate quantitative differences in expression. However, the major weakness is that only the genes already known and spotted on the arrays can be analyzed: discovery of new genes is not possible. Moreover, when using cross-species hybridization, an important proportion of the spots have to be ignored because of poor hybridization with genomic DNA, indicating nucleotide substitutions level between taxa.

The massively parallel sequencing methods, commonly called “next generation sequencing” gained in popularity approximately since 2010, due to the high number of short reads they provide, rapidly and relatively economically. These technologies enable a near complete transcriptome sequencing (i.e RNA-seq) at an affordable cost in many species without a sequenced genome (Braeutigam and Gowik 2010). It is however recognized that quantifying the difference in expression

may not be accurate in many cases, such as for polyploid taxa, for recently duplicated genes, or for genes evolving rapidly at the nucleotide sequence level (Braeutigam and Gowik 2010).

Another less commonly used method to identify differences among whole transcriptomes of different populations is the cDNA-amplified fragment length polymorphism (cDNA-AFLP), where cDNA is cut with restriction enzymes and amplified. Amplification products are then separated by electrophoresis. Bands that differ between the two transcriptomes can be cut from the gel, reamplified and sequenced. This technique has been used on *Boechera holboellii* populations (Knight et al. 2006). The main advantage of the cDNA-AFLP is that no prior knowledge on the genome is necessary. Limitations are that it can only be used among very closely related populations or lines (Vuylsteke et al. 2006), it is labor intensive, only moderate numbers of bands can be analyzed, and it does not allow quantitative estimates. Since 2010, the preferred

The suppression subtractive hybridization (SSH) method is a polymerase chain reaction (PCR)-based cDNA subtraction method able to uncover transcriptome differences in species and populations where the genome is unknown (Diatchenko et al. 1999) and is now available as a commercial kit (PCR-select kit, Clontech, Mountain View, CA).

In the suppression subtractive hybridization (SSH) method, the two different pools of cDNA being compared are digested with a blunt end four base-cutting restriction enzyme. The cDNA of interest (called tester) is divided in two tubes and each aliquot is ligated with a different adaptor. Each of the two adaptor-ligated tester cDNAs is hybridized separately with a large amount of the other pool of cDNA (called driver), to which it is being compared. Fragments in common between the two compared cDNAs (i.e. not differentially expressed), hybridize. The two volumes resulting from this first hybridization step are then combined and now only

the fragments that were unique to the tester cDNA will hybridize. This hybridization product serves as template in two rounds of PCR amplification, using adaptor primers. PCR products can be easily inserted into a vector and, after *E. coli* transformation, clones can be sequenced and analyzed. This technique is more commonly used to detect genes that changed their expression status following a stimulus in one organism. It has also been used for analyzing different strains, cultivars, genotypes, and varieties of plant species (Degenhardt et al. 2005; Wang et al. 2005; Zhang et al. 2005b; Ageorges et al. 2006; McCubbin et al. 2006). Although the amount of false positives, genes not differentially expressed but still present in the subtracted library, was considerable in some studies (Wang et al. 2005), the great advantages of suppression subtractive hybridization (SSH) are that it allows de novo genes discovery and does not require any a priori knowledge of the genome.

Table 2.2 Overview of studies reporting differentially expressed genes (DEG) among plant taxa of various evolutionary distances. The overview shows that this strategy is useful for a variety of questions and accepts diverse RNA sources. Studies are grouped according to technological platform, first with hybridization arrays followed by sequence-based methods.

Taxa compared for genes differential expression (lineages, accessions, genotypes or species)	RNA source; array used (when applicable), or sequence method.	Proportion or number of differentially expressed genes, other findings.	Functional categories or GO terms for genes differentially expressed genes between taxa.
Hybridization arrays – same species			
Five accessions of *Arabidopsis thaliana*, under standard growth condition (Chen et al. 2005).	GeneChip microarray of 8 300 genes. RNA from 10 different organs, at different plant ages.	Exclude 792 genes due to unequal hybridization in different accessions. 65 highly plastic DEG in each accession. Up to 376 plastic DEG between 8 accessions	Signal transduction, transcription and stress/defense response.
Seven *Arabidopsis thaliana* accessions, under standard growth condition or following salicylic acid application (Kliebenstein et al. 2006).	ATH1 GeneChip RNA from whole adult plant, prior to bolting.	From 1428 to 3334 DEG. Gene sequence and expression divergence are positively correlated	Responses to biotic and abiotic stimulus, stress responses, and signal transduction
Ten *Arabidopsis thaliana* ecotypes, under standard, cold or heat growth conditions (Swindell et al. 2007).	ATH1 Affymetrix microarray 22746 genes RNA from aerial parts of plantlets.	666 genes have expression that correlates with average yearly temperature. 128 (43 excluding one outlier ecotype) DEG also respond to growth temperature.	No GO terms significantly overrepresented among the 43 genes.
Five natural Chinese populations of *Arabidopsis thaliana* under standard or cold growth conditions (Engelmann et al. 2008).	ATH1 Affymetrix GeneChip, RNA from 3 weeks old plantlets, prior to bolting.	Under standard conditions, from 513 to 1482 DEG in one Chinese accession compared to Col-0. 87 DEG in all the Chinese accessions compared to Col-0	Metabolism
Two accessions of *Arabidopsis thaliana*, and their F1, grown under standard controlled condition in Petri dishes (Zhang et al. 2008).	*A. thaliana* genome tiling 1.0 F array Affymetrix of 29 409 genes, RNA from 3 days old seedlings.	1295 DEG between the two accessions, 1249 upregulated in Col-0, 676 upregulated in Van-0.	In Col-0 line: chlorophyll process, response to various stimulus, sterol process. In Van-0 line: flavonoid process, translation.

Two *Arabidopsis thaliana* accessions, from contrasting precipitation regime, under standard or mild drought growth conditions (Juenger et al. 2010).	ATH1 Affymetrix GeneChip RNA from leaves at, or prior to bolting.	512 genes not compared for expression because they differ at sequence level. 3072 genes upregulated in the drought-adapted; 2924 in the moist-adapted (5996 DEG total) under standard growth conditions. 172 of these also vary in expression after mild drought.	20% have no annotation, otherwise: Protein dynamics, transcriptional regulation, transport, hormone (auxin), light signaling, development, and DNA synthesis and chromatin structure
Five high-altitude lineages of maize, under normal and very high UV-B radiation, in the field (Casati et al. 2006).	Maize Unigene I array of 5664 cDNAs (genes). RNA from top leaves.	1947 spots differ in expression in at least one of the five high-altitude lines	In high altitude lineages: chromatin remodeling and DNA repair.
Compare two ecotypes of *Thlaspi caerulescens*, one from a site contaminated with Zn, Cd and Pb, one from a non-metalliferous soil, grown in soil different zinc content (Plessl et al. 2010).	cDNA microarrays of 4400 clones from *T. caerulescens* (709 genes) RNA from 6 weeks old adult plants, roots and shoots tissue.	168 DEG in roots between the two ecotypes, 113 in shoots (24 in common). 122 genes upregulated in Zn-rich site ecotypes, 167 downregulated genes for roots. 76 genes upregulated in Zn-rich site ecotype, 64 downregulated genes for shoots	36 unknown genes. Oxidative stress, transport processes, disease and defense mechanisms.
Hybridization arrays – cross–species			
Parental species, *Helianthus annuus* (clay soils), *H. petiolaris* (sandy soils), and the hybrid *H. deserticola* (dry sandy soils) (Lai et al. 2006).	*Helianthus* cDNA array of 2897 genes. RNA from whole plant, 6 weeks old plantlets	206 DEG between the two parents. 183 and 194 DEG between the hybrid and one or the other parent.	Transport, many other different functional classes.
Wild and weedy populations of *Helianthus annuus*, from growth chambers under standard growth conditions (Lai et al. 2008).	*Helianthus* cDNA array of 3198 genes. RNA from whole, 6 weeks old plantlets.	22 and 28 DEG between two wild and weedy comparisons. 609 DEG in at least one comparison. The different weedy populations show unique gene expression profiles.	Response to biotic and abiotic stimulus; response to stress.
The parental *Senecio chrysanthemifolius* (low altitudes) and *S. aethnensis* (high altitudes) and the hybrid *S. squalidus*; from growth chambers under standard growth conditions (Hegarty et al. 2009).	*Senecio* cDNA array of 6912 unknown spots. RNA from mature flower buds.	225 DEG (unknown cDNAs) between the parents. 311 DEG (unknown cDNAs) in the hybrid compared to the mid-parent value. Only 65 of these also differ between the parents.	Most of unknown function. Lipid binding, defense.

Three populations of *Picea sitchensis* from Alaska to California, grown outdoor in British Columbia, collected in the fall (Holliday et al. 2008).	*P. sitchensis* cDNA microarray of 21 840 genes RNA from leaves of 4 year old trees grown in common garden.	326 DEG (upregulated) in the Alaska population compared to the California, and 598 downregulated.	In Alaska populations: carbohydrate and lipid metabolism, stress response, In California population: stress response, metabolism, signal transduction.
Two *Pinus taeda* populations that differ in wood quality, harvested in summer and autumn (Yang and Loopstra 2005), grown in common garden.	*Pinus taeda* microarray of 2171 ESTs RNA from xylem from adult trees.	In latewood, 131 DEG between the two populations, in earlywood 51 DEG.	Stress response, unknown function.
The extremophile *Thellungiella salsuginea* (*halophila*) and *Arabidopsis thaliana*, grown under standard or saline growth conditions (Taji et al. 2004).	*Arabidopsis* RIKEN array (7000 genes). RNA from one month old plants.	83 DEG (upregulated) in *T. salsuginea* compared to *A. thaliana* under normal growth conditions Only 6 genes responsive to salt stress in *T. salsuginea* but 40 in *A. thaliana*.	Response to abiotic and biotic stimulus.
The extremophile *Thellungiella salsuginea* (*halophila*) and *Arabidopsis thaliana*, under standard or saline growth conditions (Gong et al. 2005).	*Arabidopsis* 'Arizona' 70mers array (25 000 genes) RNA from whole adult plant, before bolting.	20% of the probes excluded after background subtraction, normalization and filtering. 2620 DEG under standard growth conditions (93 when cutoff is threefold difference). Under saline stress, 318 genes are upregulated only in *A.thaliana*, and 128 only in *T. salsuginea* (*halophila*).	GDSL lipases ABA response, histone, nuclear ribonucleoproteins, redox control proteins. 50% of the genes responsive to salt in only *T. salsuginea* (*halophila*), are unannotated.
Arabidopsis thaliana, *A. arenosa*, and a synthetic allotetraploid. Adult plants prior to bolting, from growth chambers (Wang et al. 2006).	*A. thaliana* 70-mer microarray of 26 090 genes. RNA from rosette leaves.	2105 DEG (upregulated) in *A. thaliana*; 1818 in *A. arenosa*. Genome doubling has a minor effect on gene expression, hybridization has a large effect	In allopolyploid: hormonal regulation and cell defense and aging.
Arabidopsis halleri a zinc hyperaccumulator, *A. petraea* a non-accumulator, and F3 progenies, under normal or high Zn condition (Filatov et al. 2006). Adult plants from	*A. thaliana* ATH1-121501 Genome Array (Affymetrix) RNA from shoots and	1129 DEG (upregulated) from leaves and 721 from roots in *A. hallerii* compared to *A. petraea* under normal conditions. Under high zinc, 1228 in shoots and 772 in	No function significantly more represented. A few genes involved in transport or redox homeostasis, several hypothetical or unknown proteins.

hydroponic culture.	roots tissue separated, before bolting.	roots between species; 97 in shoots and 48 in roots between the F3 accumulator and non-accumulator progeny.	
Compare late or early flowering accessions of *Capsella bursa-pastoris*, under a vernalization or non-vernalization regime (Slotte et al. 2007). Plantlets grown in agar plates.	*A. thaliana* CATMA 25k microarrays RNA from seven-day-old seedlings, on agar MS plates	874 DEG between vernalized seedlings of the two accessions. Of the known 214 flowering time genes, 21 DEG between the two accessions.	Unknown biological process, circadian rhythm, GA metabolism and signaling.
Compare synthetic allohexaploids originating from crosses between the tetraploids *Triticum turgidum* and diploid *Aegilops tauschii*; and natural wheat (*Triticum aestivum*) allohexaploids (Chague et al. 2010)	Affymetrix GeneChip Wheat Genome Array of 55049 genes. RNA from shoots of 35 days old plants, with 5 leaves.	Approximately 5200 DEG between *T. turgidum* and *A. tauschii*. 397 DEG between the two *A. tauschii* subsp. *tauschii* and subsp. *strangulata*.	Not clearly discussed, but DEG with non-additive pattern of expression in the polyploid are enriched in photosynthesis pathways, transcription and response to stimuli.
Compare *Thlaspi caerulescens*, a zinc (Zn) hyperaccumulator, and *Thlaspi arvense* grown in agar plates or compost (Hammond et al. 2006).	Affymetrix *A. thaliana* ATH1-121501 GeneChip RNA from shoot tissue, from 64 days old plants.	3% of the genes on array excluded due to poor hybridization to *Thlaspi* gDNA. 5782 DEG (3816 higher and 1966 lower) expression) in the shoots between *T. caerulescens* and *T. arvense*.	Differential expression not presented by categories. Two plant defensin and two fatty acid desaturase genes among the most differentially expressed.
Compare *Arabidopsis thaliana* (Col-0) and *Thlaspi caerulescens* a hyperaccumulator; grown in soil with deficient, sufficient or excess zinc (van de Mortel et al. 2006).	Agilent *A. thaliana* 3 60-mer microarrays of 37683 genes. RNA from seven weeks old plantlets, roots only.	Exclude 220 genes because of poor hybridization to *T. caerulescens*. 2272 DEG (5-fold) upregulated in T. caerulescens compared to *A. thaliana*, 121 of them also responsive to zinc exposure	1147 unknown genes Response to stress (e.g. PDF); metal homeostasis genes (ZIP gene family), lignin biosynthesis
Populus trichocarpa and *Arabidopsis thaliana*. *P. trichocarpa* adult grown in greenhouse *A. thaliana* from public databases. Compare 4188 orthologous genes (Quesada et al. 2008).	P. *trichocarpa* array 60-mere of 42364 genes. RNA from 6 different organs of *P. trichocarpa*.	In roots and young leaves 40% DEG. For genes expressed in many organs, expression is conserved among species.	Regulation of gene expression
Fourteen Brassicaceae taxa, grown in chambers in MS-agar Petri plates. 19 days old rosette leaves (Broadley et al. 2008).	*A. thaliana* ATH1-121501 GeneChips RNA from leaves of adult plants.	18494 genes (out of 22756) analyzed, the others 4252 (18.7%) are excluded because of unequal hybridization. Transcriptome divergence correlates with	Not presented.

		evolutionary distance.	
Compared alpine *Pachycladon fastigiata* from mid-altitudes and *P. enysii* from high altitude (Voelckel et al. 2008).	*A. thaliana* AROS version 1.0 genome set (Operon Biotechnologies), RNA from field-collected leaves	24% of genes excluded because of poor hybridization. 310 genes are upregulated in *P. fastigiata* and 324 in *P. enysii* at 1.5-fold.	In mid-altitude: Stress and hormone response; In high-altitude: Cell wall and translation.
Compare the *Spartina maritima*, *S. alterniflora*, the hybrid *S.* X *townsendii* and its allopolyploid *S. anglica*. Adults plants, transplanted from the field to growth chambers (Chelaifa et al. 2010b).	44 K rice array (Agilent G2519F) of 21 509 genes RNA from leaves	30% of the genes excluded because absence of hybridization signal. 1235 DEG between the hybrid *S.* X *townsendii* and the paternal parent. 497 DEG between the allopolyploid *S. anglica* and the hybrid	Hybrid and parents: genes with unknown function, carbohydrate metabolism. Hybrid and allopolyploid: development and cellular growth, transport
Compare the salt-marsh sister lineages *Spartina maritima* (declining) and *S. alterniflora* (invasive), Adults plants, transplanted from the field to growth chambers. (Chelaifa et al. 2010a).	44 K rice array (Agilent G2519F) of 21 509 genes RNA from leaves.	30% of genes on array cannot be used because of absence of signal. 1247 DEG between the parents *S. maritima* and *S. alterniflora*.	Development and cellular growth, metabolism of proteins, carbohydrate and lipid systems.
Sequence-based methods			
Compare three genotypes of *Eragrostis curvula*: a diplosporous tetraploid, a sexual diploid and an artificial sexual tetraploid (Cervigni et al. 2008).	RNA from inflorescence, in a pre-meiotic developmental stage. Sequences from untreated cDNA libraries	The 12 600 randomly selected clones assemble into 8824 unigenes. From that, 112 are DEG in at least one comparison.	Not discussed
Two populations of *Boechera holboellii* from contrasting water availability environments. Grown under moist or dry conditions Knight et al. 2006).	RNA from leaves of 5-month-old plants. cDNA-AFLP	Sequenced 300 of the 450 cDNA-AFLP fragments different between drought and control. Discussion on 24 genes.	Signal transduction, transcription, redox regulation, oxidative stress and stress
Genes enriched after heat stress in two *Festuca sp.* with contrasting heat tolerance.	RNA from shoots and leaves of adult plants.	Of 2495 ESTs analyzed, 1800 clustered into 1090 genes (434 contigs and 656 singlets).	More than 30% are novel genes. Heat-tolerant: cell maintenance,

Grown in chambers under normal and heat stress (Zhang et al. 2005b).	Library subtraction PCR-select		photosynthesis, protein synthesis, signaling, and transcription factor. Heat-sensitive: metabolism and stress.
Drought stress-resistant and stress-susceptible rice genotypes. 5 weeks old plants, in growth chambers. Examine subtracted libraries enriched for genes induced by drought in the two genotypes (Wang et al. 2007).	RNA from whole plants, roots and shoots tissues separately. Library subtraction PCR-select	From 2112 clones, 1991 have single insert, 7% of roots and 13% of shoots are truly DEG, as validated by cDNA microarrays. They represent 39 and 31 genes.	6 unknown genes In resistant: signal transduction. In susceptible: metabolism, ribosomes, Response to drought: Transcription, protection against oxidative stress.
Two accessions of *Medicago truncatula* with contrasting responses to ozone. Examine subtracted libraries enriched for genes responding to ozone of 8 weeks old plants in one sensitive, and in one resistant accession (Puckette et al. 2009).	RNA from leaves, at different time point after ozone treatment. Library subtraction PCR-select	From 2500 clones, 800 are sequenced, that assemble into 239 unique DEG. Of the DEG, 183 genes are from the susceptible, 56 genes from the resistant	15 novel genes. In resistant: response to stress genes In susceptible: oxidative stress, cell growth, and translation
Two *Phaseolus vulgaris* varieties with contrasting tolerance to drought, grown in greenhouse, under standard or drought condition (Montalvo-Hernandez et al. 2008).	RNA from leaves and roots tissues of 40 days old plants. Library subtraction PCR-select	500 clones sequenced, 18 DEG after stringent criteria on macroarrays	6 genes similar have unknown function, otherwise: Stress, defense, signal transduction. Aquaporin
Two *Malus domestica* cultivars with contrasting sensitivity to apple scab (*Venturia inaequalis*), uninfected (Degenhardt et al. 2005).	RNA from leaves of adult trees, in greenhouse. Library subtraction PCR-select	262 in the resistant (total 480 clones). 218 clones in the susceptible Less than 10% false positive after stringent criteria on macroarrays	23 novel genes. In resistant: Defense, metallothioneins. In the susceptible: photosynthesis,
Two rice (*Oryza sativa*) lines one susceptible, one resistant to brown planthopper insect pest (*Nilaparvata lugens*) (Wang et al. 2005). Plants harvested at different time points after infestation.	RNA from young and mature leaves. Library subtraction PCR-select	From 5700 clones isolated, 154 clones represent 136 unique truly DEG after stringent criteria of macroarrays screening and sequencing	Almost 50 % are of unknown function and nine novel genes. Enriched in resistant rice: Disease and stress, signal transduction, electron transport.
Two grape (*Vitis vinifera*) cultivars with different colors of berries pulp, red or white. Berries harvested from field grown plants	RNA from grapes berries pulp; pedicel, seeds, and peel	1600 selected cDNA clones give 1406 ESTs that assemble into 96 genes.	19 novel genes, 7 of unknown function. Red pulp: Secondary metabolism (including anthocyanin biosynthesis), energy

(Ageorges et al. 2006).	removed. Library subtraction PCR-select		White pulp: Stress, unclassified.
Two wheat (*Triticum aestivum* L.) lines, one susceptible and one resistant to head blight *F. graminearum*. After inoculation, grown in chambers (Bernardo et al. 2007).	RNA from inflorescence, at different time point after inoculation. Library subtraction PCR-select	From 2306 cDNA clones, 44 are unique truly DEG after stringent criteria of macroarrays screening and sequencing.	16 novel genes; 6 unknown proteins In resistant: defense, In susceptible: several genes, diverse function.
Geothermal *Agrostis scabara* and heat sensitive *A. stolonifera* grass species, grown in chambers, under standard or heat conditions (Xu et al. 2008).	RNA from leaves and roots of adult 8 weeks old plants. Differential display	From 63 polymorphic fragments, 26 were sequenced. Find six genes	8 novel genes In resistant: expansin

Note: To illustrate the state of knowledge in the community, at the time our research on comparative transcriptomics of *Oxytropis* was performed; only the papers published earlier than 2011 are reported in this table.

The several examples listed above (Table 2.2) illustrate that the strategy of whole transcriptome comparison between taxa can reveal potentially new adaptive features and generate novel hypotheses. Often very closely related taxa were compared, differing only in a minor feature, for instance resistance or susceptibility to a particular stress. Species not closely related can also be directly compared for genome-wide differences in gene expression, as shown by the 14 different taxa of the Brassicaceae family hybridized to the Affymetrix ATH1 array (Broadley et al. 2008). In these cases, though, an important fraction of the genes spotted on the array, sometimes up to 30% (Broadley et al. 2008; Voelckel et al. 2008; Chelaifa et al. 2010a), cannot be used because nucleotide sequence divergence between taxa results in unequal hybridization that is not due to difference in expression.

The plant growth conditions, tissues, or treatments compared differ widely among the studies, illustrating the flexibility of the comparative transcriptomics strategy. Most often, RNA was extracted from leaves of plants grown in controlled conditions, especially in earlier studies presumably for the simplicity of this tissue, but whole plants, different tissue or very small plant organs are now commonly being compared. Field grown plants are also directly examined for gene expression, in an effort to represent more realistic growth conditions (Yang and Loopstra 2005; Holliday et al. 2008; Travers et al. 2010). The comparative transcriptomics strategy is also scalable, genome-wide expression can consist in a comparison between two taxa, or the dataset may include several different organs, and treatments compared among several species.

As mentioned previously, hybridization arrays do not allow novel gene discovery, nevertheless this strategy is able to identify differentially expressed genes of unknown function that were already spotted on the array. On the contrary, novel genes are almost always present using library subtraction techniques, even in very

well studied plants, such as rice (Wang et al. 2005; Wang et al. 2007) or *Medicago truncatula* (Puckette et al. 2009).

Differential gene expression profiles are now commonly presented in terms of gene functions overrepresented in one or the other gene library. Common tools for this purpose are the Gene Ontology terms (Ashburner et al. 2000) and the MIPS functional categories for plant biology (Ruepp et al. 2004). The functions overrepresented in sets of differentially expressed genes are diverse according to the above examples (Table 1.2), and this illustrates the usefulness of the approach to gain novel and untargeted views. One surprising finding from several studies listed is the recurrent differential expression of genes related to stress, defense or aging between taxa that do not differ in stress tolerance (Chen et al. 2005; Yang and Loopstra 2005; Ageorges et al. 2006; Wang et al. 2006; Lai et al. 2008), or between taxa from latitudinal or altitudinal gradients (Holliday et al. 2008; Voelckel et al. 2008; Hegarty et al. 2009).

Genes will experience a change in their expression profile through evolutionary time, not only under selective pressure, but also due to selectively neutral stochastic processes. In all comparative transcriptomics studies, a major challenge resides in distinguishing the genes involved in adaptation among the selectively neutral differences. To aid in this search, gene expression profiling before and after application of the stress under study (e.g. cold, heat, UV, ozone, fungus, insect) is examined in many cases. In conclusion, the field of comparative transcriptomics among taxa has both the flexibility and the feasibility to generate novel hypotheses on the complex question of plant adaptation to the rigorous arctic climate.

2.1.4.2 Sequence divergence in coding sequences can also suggest candidate evolutionary relevant features

The functional roles of a protein region impose specific selective pressures on its sequence evolution (Kosakovsky Pond and Frost 2005b; Nielsen 2005; Anisimova and Liberles 2007): negative selective pressure (negative selection) removes deleterious alleles and is detected when nonsynonymous mutations (d_N, that alter the encoded amino acid) are significantly less frequent than synonymous ones (d_S, that code for the same amino acid). Positive selection is inferred when nonsynonymous mutations *(d_N)* are more frequent than synonymous ones *(d_S)* and indicates that novelties were once advantageous at the protein sequence level. Natural selection has therefore shaped the pattern of coding sequence variation between related protein-coding genes; and statistical tools now exist to detect and identify location of particular codons that evolved under negative or positive selection (Kosakovsky Pond et al. 2005; Anisimova and Liberles 2007). Divergence based computational tools (Kosakovsky Pond et al. 2005; Nielsen 2005) are extremely easily applicable to a wide array of questions, because they do not require assumptions about population demography. The only requirement is that analyzed sequences should not have experienced important recombination (Anisimova and Liberles 2007).

Detecting positive selection is particularly appealing because it is associated with adaptation and the evolution of new form or function (Nielsen 2005). In an evolutionary context, tools for codon selection detections were applied to narrow down a subset of genes that may carry important biological novelties among large set of related sequences, such as sequences generated by ESTs or genome sequencing initiatives, from various kind of organisms (Mondragon-Palomino et al. 2009; van der Aa et al. 2009; Zamora et al. 2009; Aguileta et al. 2010; Elmer et al.

2010). In a few cases, the particular sites under positive selection were further functionally tested to confirm that the sites identified truly carry a biological role. For instance, the evolution of substrate specialization of salicylic acid methyltransferase (SAMT), a protein involved in formation of floral scent compounds or in cellular detoxification was analyzed and only one codon was detected as under positive selection. Further functional studies (site-directed mutagenesis, protein expression and enzymatic assays) confirmed that ability of SAMT to discriminate substrate is due to presence of methionine instead of histidine at that positively selected codon (Barkman et al. 2007). This substrate discrimination has impacts on the plant secondary metabolism and in floral scent, which are important ecological and evolutive traits in plants. Another similar study detected a few positively selected sites in the plant resistance gene eukaryotic translation initiation factor 4E (*eIF4E*) in a set of susceptible and resistant alleles from pepper, tomato and pea (Cavatorta et al. 2008). The positively selected sites cluster in the protein region that contains resistance determining sites and are known to interact with viral pathogens such as *Potato virus Y* (PVY) and *Tobacco etch virus*. In another example, evolution of the phospholipase A2 (PLA_2) gene sequences, a venom-related gene, was analyzed from fours species of *Sistrurus* rattlesnakes, which feed on different preys. Nine amino acids residues under strong positive selection were identified, with a disproportionately high proportion of these on the surface and in the anticoagulant functional regions of the PLA2 protein (Gibbs and Rossiter 2008) suggesting that positive selection has led to high levels of functional diversity in proteins among these snakes. The few above examples illustrate that the methods for detecting positive selection are now sensitive enough to reveal codons substitutions with genuine phenotypic effects.

Earlier calculations estimated the averaged d_N:d_S ratio over the entire gene coding sequence, and required strong selection acting on the entire gene to allow

detection, but current methods have now improved sensitivity and are able to detect selection on a site-by-site basis (Kosakovsky Pond et al. 2005; Anisimova and Liberles 2007). Three methods for estimating selective pressure on a coding sequence are implemented in the HyPhy package (Kosakovsky Pond et al. 2005). The SLAC (Single Likelihood Ancestor Counting) is a counting method that involves reconstruction of the ancestral sequence and is more conservative than two other methods. The REL (Random Effects Likelihood) is computationally complex, and involves fitting a distribution of rates across sites and then, rates for individual sites. The FEL (Fixed Effects Likelihood) involves fitting a distribution on a site-by-site basis, with no assumption of the distribution of the rates across sites. The three methods converge on identifying sites under positive or negative selection when at least eight sequences are analyzed (Kosakovsky Pond and Frost 2005b).

In the present thesis, codons that evolved under negative and positive selection are identified in sequences of gene families' orthologs and paralogs and in low copy genes for different *Oxytropis* species, for genes that exhibit contrasting expression between arctic and temperate *Oxytropis* species.

Chapter 3

3 Connecting text to Chapter 3

This chapter compares subtracted transcriptome composition (expressed genes) between plantlets of two arctic and two temperate *Oxytropis* species, as a way to discover yet unknown molecular specializations of arctic plants. This is, to our knowledge, the first description of a true arctic plant transcriptome. When the study was initiated, massively parallel sequencing or next generation sequencing (NGS) methods were not yet widely accessible, and cDNA library subtraction followed by medium-scale clone sequencing was a standard tool for gene discovery of non-model organisms. The molecular techniques involved in this chapter include cDNA library subtraction (suppression subtractive hybridization, SSH) from plantlet total RNA and clone screening and sequencing (ESTs). Sequence analyses were semi-automated for annotation and comparison to public sequence databases. The set of genes exclusively transcribed in arctic *Oxytropis* plantlets reveals categories of genes that were not previously suspected to participate in arctic adaptations, and confirm previously known responses of cold-acclimating plants. Transcript levels of selected genes were characterized in different *Oxytropis* species grown under the two growth conditions, arctic and temperate, using real-time RT-PCR to support the library subtraction results. The real-time RT-PCR data also highlights that the expression profile of one gene could be driven by adaptation to arctic climate, but that expression profiles of other genes are mainly species-specific, rather than environment-specific.

A highly condensed version of this chapter was published in Functional and Integrative Genomics in April 2011 (DOI: 10.1007/s10142-011-0223-6), with “Arctic and temperate plant gene expression” as running title and “arctic, plant, gene expression, *Oxytropis*, library subtraction, defense response” as keywords. The

article title was “PR-10, defensin and cold dehydrin genes are among those over expressed in *Oxytropis* (Fabaceae) species adapted to the Arctic” (Archambault and Strömvik 2011). The longer version is presented in this chapter with kind permission from Springer under license number 2695381485945; it explains methodology in more details and includes measures of pairwise divergence for background clones (potential false-positives), and for similar unique genes.

3.1 PR-10, defensin and cold dehydrin genes are among those over expressed in *Oxytropis* (Fabaceae) species adapted to the Arctic

3.1.1 Abstract

In many studied plants, typical responses to cold treatment include up-regulating the hydrophilic COR/LEA genes and down-regulating photosynthesis related genes, carbohydrate metabolism, GDSL-motif lipase, hormone metabolism and oxidative regulation genes. However, next to nothing is known about gene expression in arctic plants, which are actually actively growing in a harsh, cold environment. The molecular mechanisms behind the many specific adaptations of arctic plants, such as slow growth, well-developed root systems and short stature, are not well understood. In this study we examine whole plantlet transcriptome differences between two arctic and two temperate *Oxytropis* (Fabaceae) species, grown under their respective controlled environmental conditions. Gene expression differences are analyzed using cDNA library subtraction followed by EST (Expressed Sequence Tags) sequencing and annotation. Sequences from a total of nearly 2000 clones cluster into 121 and 368 unique genes from the arctic and from the temperate plants, respectively. A set of novel genes forms the core of the arctic-enriched gene library, and the predominant biological process for genes from this library is “response to stimulus”. A concurrent overrepresentation of PR-10, plant defensin and KS-dehydrin genes in the transcriptome is a novel feature for species adapted to stressful growth environment. The temperate-enriched genes are involved in photosynthesis, translation and nucleosome assembly. Interestingly, both arctic and temperate-enriched libraries also contain genes involved in ribosome biogenesis and assembly, however of different types. Real-time reverse transcription

PCR of KS-dehydrin and two PR-10 genes, as well as the light harvesting complex b1 gene supports the library subtraction data.

3.2 Introduction

The challenges to arctic plant life go beyond the severe winter temperatures - very short growing season, common summer frosts, strong winds and low light quality are limiting conditions to plant growth. These tough little plants present a suite of morphological and physiological specializations compared to their temperate relatives. They have long life cycles (Grulke and Bliss 1988), leaves or flower primordium development extend on for many growth seasons (Sørensen 1941), and they sport well-developed root systems (Bliss and Gold 1999). Furthermore, their photosynthetic and respiratory apparatus is more efficient at 10 °C than at higher temperatures (Pyankov 1991; Xiong et al. 1999; Semikhatova et al. 2007), and they can tolerate freezing temperatures while still actively growing (Junttila and Robberecht 1993). However, the understanding of the molecular mechanisms behind these adaptations is at best fragmentary.

Because of the potential adaptive value of variation in gene expression (Whitehead and Crawford 2006b; Whitehead and Crawford 2006a; Swindell et al. 2007), we explored differentially expressed genes between arctic and temperate plants. We chose the *Oxytropis* (Fabaceae) genus as our model system. *Oxytropis* is predominantly distributed in the temperate and boreal regions of North America and Northern Asia, and also includes 44 arctic species (Elven 2007). Although only eight species occur in the Canadian Arctic, this is the legume genus with highest species diversity in that area (Aiken et al. 2007). We used the suppression subtractive hybridization (SSH) technique to discover differences in the

transcriptomes of two arctic *Oxytropis* species and two temperate *Oxytropis* species. An important advantage of this technique is that it is untargeted and enables the discovery of potentially relevant, novel genes from uncharacterized genomes. In a SSH experiment, the hybridization steps render genes similarly abundant in both cDNA libraries compared unavailable for the subsequent PCR amplification step (Diatchenko et al. 1996). As a result, genes differentially expressed can be amplified exponentially. Despite potential biases that may arise because PCR amplification efficiency may differ between transcripts, the EST coverage for a unique gene in the final subtracted library should roughly reflects the relative transcript fold differences between the two libraries compared. EST coverage is not expected to reflect absolute abundance in the original transcriptome. However, some differentially expressed genes may be absent from the final subtracted library. Real-time reverse transcription PCR that can provide a more accurate indication of the relative abundance of a gene in the original mRNA populations, supports findings of the SSH study for four selected genes. This study is the first description of an arctic plant transcriptome and we present several important differences in gene expression between the species.

3.3 Materials and methods

3.3.1 Plant material and RNA extractions

Seeds of the arctic species *O. maydelliana* and *O. arctobia* and the temperate species *O. splendens* and *O. campestris* subsp. *johannensis* were scarified, sterilized and stratified at 4°C on 1/2 MS Basal Medium (Sigma, Oakville, Ontario) agar plates. The seed sources are listed in Supplementary Table S3.1. Germinating seeds

were placed in growth chambers mimicking the summer-fall conditions in temperate climates (16 h of light of 225 μmol/m^2/s at 22˚C, and 8 h darkness at 18˚C) or in the low arctic (20 h of light of 150 μmol/m^2/s at 10˚C, and 4 h darkness at 10˚C). RNA was extracted (Qiagen) from plantlets at the two leaf stage.

3.3.2 Suppressive subtraction cDNA library construction

Extracted RNA from two plantlets of a species were pooled prior to cDNA synthesis, and one microgram of this pool was used as template for cDNA synthesis (Table 3.1) using the SuperSmart PCR cDNA Synthesis kit (Clontech, Mountain View, CA) Genes present uniquely in either the arctic or in the temperate species plantlet transcriptome were isolated by applying the suppression subtractive hybridization strategy (Diatchenko et al. 1996) using the commercial PCR-select cDNA subtraction kit (Clontech, Mountain View, CA). The subtraction compared a pool of plantlet cDNA from the two arctic species (*O. arctobia* and *O. maydelliana*) grown in simulated arctic environment, and a pool of plantlet cDNA from the two temperate species (*O. campestris* subsp. *johannensis* and *O. splendens*) grown in simulated temperate environment. The subtraction was performed in both directions resulting in two subtracted libraries: one “arctic-enriched” library and one “temperate-enriched” library (Table 3.1). As recommended by the manufacturer (Clontech manual PT-1117), the first hybridization involved a 1 to 5 ratio of tester to driver cDNA quantity, while the second hybridization involved an additional ratio of 1 to 3.3 of tester to driver. Hybridizations were followed by the recommended 27 cycles of primary PCR and 12 cycles of secondary PCR. As recommended by the manufacturer, a PCR analysis of subtraction efficiency was also performed.

Table 3.1 Starting material for construction of *Oxytropis* arctic-enriched and temperate-enriched libraries with PCR-Select cDNA subtraction protocol.

RNA material	Ploidy (2n=)	Tester cDNA	Driver cDNA	Resulting clones
Two plantlets of *O. arctobia* (grown in arctic conditions) Two plantlets of *O. maydelliana* (grown in arctic conditions)	2x=16 12x=96	Pool of *O. arctobia* cDNA and *O. maydelliana* cDNA	Pool of *O. c. johannensis* cDNA and *O. splendens* cDNA	"arctic enriched" Oapa01 library (forward library)
Two plantlets of *O. c. johannensis* (grown in temperate conditions) Two plantlets of *O splendens* (grown in temperate conditions)	6x=48 2x=16	Pool of *O. c. johannensis* cDNA and *O. splendens* cDNA	Pool of *O. arctobia* cDNA and *O. maydelliana* cDNA	"temperate enriched" Otpt01 library (reverse library)

PCR products of arctic and temperate subtracted libraries were cloned non-directionally using a TOPO TA kit and transformed into ElectroMAX DH10B electro competent cells (Invitrogen, Carlsbad, California). White colonies were grown in a total of thirty 96-well plates containing SOC-ampicillin liquid medium, and subsequently screened by PCR using vector primers.

3.3.3 Sequence analysis and annotation of subtracted library clones

Clones with confirmed single inserts were sequenced single-pass (McGill University and Génome Québec Innovation Center). Sequences were basecalled using Phred (Green 2002) and trimmed using SeqTrim (Falgueras et al. 2007). The 1114 arctic and 613 temperate processed EST sequences were assembled into

contigs using Phrap (Green 2002) with the following parameters: minmatch 50; minscore 100, indexword 10; qual_score 20, as described (Vodkin et al. 2004; Strömvik et al. 2006). As a first step, the Phrap algorithm automatically discards exact duplicate sequences that have the same sequence and same length. Sequences with homopolymers repeats within the set of singlets sequences were discarded. Images of contigs assemblies were visually inspected to mark suspected chimera of misassembled sequences. The 693 sequences from arctic species and 552 sequences from temperate species that are 200 bp and longer appear in public sequence databases, including the duplicate sequences discarded by Phrap assembly (Table 3.2), under the GenBank accession numbers GW696871 to GW698115, while the 406 arctic and the 4 temperate sequences less than 200 bp in length are available upon request to the author. For each subtracted library, the contig sequences, singleton contigs and singlets ESTs were combined in one set of sequences that represent the unique genes. Blast2GO (Conesa et al. 2005; Gotz et al. 2008) was used to annotate (in May 2009) unique genes, considering up to 25 similar hits of the NCBI non-redundant database (nr) at the permissive 1 e-05 E value cutoff using the BLASTX program. The effective E value was, however, always smaller than 2 e-06, as seen in the results section. Blast2GO was also used to assign Gene Ontology terms (Ashburner et al. 2000) to unique genes.

Oxytropis unique gene sequences were manually classified into general categories considering the assigned GO (Gene Ontology) terms for biological process, the MIPS functional categories (Ruepp et al. 2004), and the similarity (by BLASTX) to the complete *Arabidopsis* peptide sequence collection (TAIR8_pep_20080412). Annotations were manually verified and curated, based on information in the Kyoto Encyclopedia of Genes and Genomes (Kanehisa and Goto 2000), the Arabidopsis Information Resource (Swarbreck et al. 2008), the Plant

Metabolic Network database (Zhang et al. 2005a) or the NCBI Entrez Gene (Maglott et al. 2005).

3.3.4 Sequence similarity among unique genes for detection of background clones and of gene copies

According to the manufacturer (Clontech manual PT-1117), a certain level of background is expected in the subtracted libraries. Background clones correspond to cDNAs found in a subtracted library, but that were from mRNAs present in similar abundance in both compared RNA samples. According to the manufacturer' manual, background may be high when the cDNA samples compared differ by only a few differentially expressed genes and have low quantitative abundance differences. By contrast, given that the cDNA libraries compared here are from different *Oxytropis* species that grew under very different conditions, only a low level of background was expected. Here, background was detected by a sequence similarity search between the genes from the two subtracted libraries. Similarity was assessed by building a local BLAST target database composed of the totality of genes from the arctic-enriched and temperate-enriched libraries; and evaluating sequence similarities to the two libraries using BLASTN and TBLASTX (Altschul et al. 1997) at the 1 e-04 E value cutoff. This permissive threshold may identify genes that are only modestly similar. The similar genes that were from different libraries were marked as potential false positives (indicated by an asterisk in Supplementary Tables S3.3 and S3.4 and listed in Table 3.4), and may correspond to background. Since each subtracted library is constituted by two different genomes, similar genes from one library may be either potential orthologous genes or paralogous genes. All similar genes were aligned using the Geneious alignment tool or the MAFFT algorithm (Katoh et al. 2002) from the Geneious program, and pairwise divergence

between each pair was calculated in MEGA (Kumar et al. 2008). Percentage pairwise divergence is calculated by the number of uncorrected base differences per site between sequences, for each pair of sequences.

3.3.5 Real-time reverse transcription PCR

Relative transcripts levels were measured with real-time reverse transcription PCR (RT-PCR) from cDNA, for four genes (dehydrin, two genes of the PR-10 family, lhcbI) displaying important difference in transcript abundance between the arctic and the temperate cDNA libraries. Specific primers (Supplementary Table S3.2) were designed in Geneious (Drummond et al. 2008) and manually adjusted to ensure that a primer anneals to a region that is conserved among the four species but that varies among the copies of a gene family. The different real-time RT-PCR products from gDNA were first sequenced for plantlets of the four *Oxytropis* species, to confirm that the gene is present in all four genomes, and that the primer pairs amplify a single product (Supplementary Table S3.2 for GenBank accession numbers). Plantlets from all four species (arctic and temperate) were grown from seeds in temperate and in arctic conditions (described above), RNA was extracted and two micrograms of RNA from single plantlets were reverse transcribed into cDNA using the QuantiTect kit (Qiagen, Mississauga).

Real-time RT-PCR reactions were performed using Brilliant SYBR Green dye (Supplementary Table S3.2) with two technical replicates per samples for each of three biological replicates. Although a higher number of biological replicates would have been desirable, only three biological replicates were measured due to limited availability of seed material. Data were analyzed with MxP3000 4.01

software (Stratagene). Normalization was carried out relative to actin gene expression (Simon 2003), with the "mean normalized expression" formula number 2.

3.4 Results

3.4.1 EST sequencing from subtracted libraries reveals different biological processes in arctic and temperate plants

In order to characterize potential adaptive differences in gene expression, we investigated transcriptome differences between arctic and temperate *Oxytropis* species. One arctic-enriched and one temperate-enriched subtracted library (by SSH) was constructed from *Oxytropis* plantlet mRNA. The subset of ESTs included in each of the steps that lead to building a list of unique genes is presented (Table 3.2). Phrap identified 253 arctic and 12 temperate ESTs as exact duplicates (same length and same sequence) that were automatically discarded from the assembly step, but that were deposited in GenBank (if 200 bp or longer). In addition, 17 arctic and 57 temperate ESTs with internal homopolymer repeats were discarded. A very few sequences were suspected to be chimera or misassembled (Table 3.2) after visual inspection of contigs images of an initial test assembly, and chimeras were excluded from the final assembly. From the arctic-enriched library, 846 of the 1112 trimmed ESTs (ESTs mean length of 605 bp) were retained, and assembled into 117 arctic unique genes (contigs, singleton contigs and singlets). From the temperate-enriched library 545 of the 613 trimmed ESTs (mean length of 809 bp) were retained, and assembled into 364 temperate unique genes (contigs, singleton contigs and singlets). Altogether, only six sequences are possibly misassembled, but do not appear to negatively affect the consensus sequences (Supplementary Figure S3.1).

To retain as much as information as possible, they were added to the set of unique genes, to build the final list 121 arctic unique genes and 368 temperate unique genes to be annotated.

Table 3.2 Subsets of ESTs retained in each step of assembly and annotation of unique genes from arctic-enriched and temperate-enriched *Oxytropis* plantlets cDNA libraries.

Category of ESTs	Number of ESTs in arctic-enriched library and file name [a]		Number of ESTs in temperate-enriched library and file name [a]		Notes
Trimmed and cleaned ESTs	1114		613		Is the SeqTrim output
ESTs suspected chimeras	2	A2_liste_arctic_suspect_chimera_2seq.txt	0		Identified after an initial assembly. Suspected chimeras not included in the final assembly. Are in GenBank.
ESTs suspected misassemblies	2	A3_liste_arctic_suspect_misassembly_2seq.txt	4	T3_liste_temperate_misassembly_4seq.txt	Are in GenBank.
ESTs used for assembly	1112	A4_liste_arctic.fasta.txt	613	T4_liste_temperate.fasta.txt	Exclude the suspected chimeras. Are in GenBank if 200b or longer.
Exact duplicates	253	A5_liste_arctic_exact_duplicates.txt	12	T5_liste_temperate_exact_duplicates.txt	Are in GenBank if 200b or longer.
ESTs with homopolymers repeats	17	A6_liste_arctic_discarded_17seq.txt	57	T6_liste_temperate_discarded_57seqs.txt	Not initially excluded by SeqTrim. Used for assembly and if output as singlets, manually excluded from final list of unique genes. Not in GenBank, except for two arctic sequences..
ESTs used in the final list of unique genes	846	Supplementary Table S3.3	545	Supplementary Table S3.4	Includes the suspected misassemblies and chimeras. Excludes exact duplicates.
ESTs 200 bp or longer	693	A8_arctic_outputfile_final	552	T8_temperate_outputfile_fi	Are in GenBank.

		.txt		nal.txt	
ESTs shorter than 200 bp	406	A9_arctic_outputfile_lesst han200.txt	4	T9_temperate_outputfile_le ssthan200.txt	Not in GenBank.
ESTs included in a in contigs		A10_liste_arcticContig.co mponents.txt		T10_liste_temperateContig. components.txt	Are in GenBank if 200b or longer.

[a] Files found in the electronic appendix CD

The compositions of the contigs are listed in Supplementary Table S3.3 (arctic) and S3.4 (temperate). The unique gene sequences were first given a putative annotation by searching similarities against public protein sequences database (Conesa et al. 2005; Gotz et al. 2008). This annotation was then manually verified, curated and assigned to general categories using several bioinformatics databases. The breakdown of the genes in functional categories is presented in Table 3.3, while the full list of annotated sequences are listed in Supplementary Table S3.3 (arctic) and S3.4 (temperate).

All techniques have limitations, and because the library subtraction (PCR-select) is an enrichment method, it is sensitive to factors such as amplification efficiency. EST coverage in the resulting sequence assembly does therefore not strictly reflects transcript proportion in the mRNA populations compared, and not all genes involved in arctic adaptations may be identified. As a result, Table 3.3 should not be interpreted as a traditional expression profiling experiment. More precisely, a high EST coverage for a gene in the subtracted library suggests it was highly differentially expressed in the original libraries compared (Diatchenko et al. 1996), but does not imply it was highly abundant in the initial mRNA population (tester, Table 3.1). Similarly, a low EST coverage for a gene in a subtracted library is a weak indication that it is differentially expressed, and many single ESTs listed in Table 3.3 are not discussed for that reason.

Sets of overrepresented genes in a sample are commonly analyzed by enrichment in Gene Ontology (GO) terms by comparison to a reference set of genes (e.g. Blüthgen et al. 2005) . Because a proper reference set of genes, such as an *Oxytropis* genome, does not yet exist, the list of *Oxytropis* genes from the enriched libraries are unfortunately not amenable to this analysis. Keeping the limitation of this study system in mind, we here begin the description of the genes potentially relevant in arctic plants. The individual genes with highest coverage in

the arctic-enriched library are listed first, followed by relevant gene categories in each library.

Most *Oxytropis* unique genes received an annotation, except for 25.6% (31/121) of the arctic-enriched genes (58.5% of the ESTs) and 7.6% (26/368) of the temperate-enriched unique genes (5.8% of the ESTs), which have no similarity to any sequences in the public database. The finding of a very high proportion of ESTs that represent novel genes in the *Oxytropis* arctic-enriched library (Table 3.3, Supplementary Table S3.3) is a striking result. Despite the permissive threshold set in the annotation assignment step, the *Oxytropis* genes received annotations based on high similarity to public known sequences, and reached values as low as 1.57 E-116 (60S ribosomal protein L8 in the arctic-enriched library, Oapa1_0051_HFY327_047.ab1) and 3.04 E-146 (LHCB4 gene of the temperate-enriched library, temperate.fasta.screen.Contig45). Only fewer that nine genes in each subtracted libraries received annotation based on E values higher than 9 E-08. The highest E value is 1.81 E-06 in the arctic-enriched library, and 9.96 E-06 in the temperate enriched-library, both for dehydrin related genes (arctic.fasta.screen.Contig25 and Otpt1_0864_CH0101_025.ab1, Supplementary Tables S3.3 and S3.4).

The majority of the ESTs from the arctic-enriched library (610 ESTs from a total of 846 unique ESTs) assembled into contigs representing seven genes. Two unique novel genes received a particularly high coverage, arctic.contig63 and arctic.contig62, with 236 and 94 ESTs, respectively. Another well-covered unique gene from that library (arctic.contig60 with 70 ESTs) was similar to an undescribed gene. Other well-covered unique genes from that library (arctic.contig61 and arctic.contig59, with 94 and 45 ESTs, respectively) were members of the “response to stimulus” category (Table 3.3 and Supplementary Table S3.3). One of them (arctic.contig61) is a pathogen-related class 10 gene

(PR-10) similar to a temperate-enriched gene (3.89 % pairwise divergence, Table 3.4) having only a shallow EST coverage (temperate.contig54 of 2 ESTs, Supplementary Table S3.4). The other stimulus response arctic gene (arctic.contig59) is a KS-dehydrin, with a weak similarity (37.71 % pairwise divergence, Table 3.4) to a unique gene from the temperate-enriched library (temperate.contig22). Two well-covered unique genes from the arctic-enriched library (arctic.contig57 and arctic.contig58, with 30 and 32 ESTs, respectively, Supplementary Table S3.3) completed this list of seven highly covered genes. They were both annotated as acetyl-carboxylase carboxyltransferase beta subunit (accD), a gene of the lipid category (Table 3.3), and were similar to a unique gene (temperate.contig41) from the temperate-enriched library (2.88 % and 5.83 % pairwise divergence, respectively, Table 3.4). In addition, 8 different defensin genes from the arctic-enriched library were composed of a total of 24 different ESTs (Supplementary Table S3.3). Overall, the arctic-enriched library is proportionally rich in novel genes and in stimulus response genes (Table 3.3).

The cDNA library from *Oxytropis* plantlets growing in temperate conditions was greatly enriched in genes of the energy and photosynthesis category (Table 3.3). The six unique genes with highest coverage were in this category (Supplementary Table S3.4). They are psaN that functions as photosystem I reaction centre subunit precursor (temperate.Contig120, 9 ESTs); *lhcbI*, a light harvesting complex of photosystem II (temperate.Contig119, 8 ESTs); *rbcS*, the small subunit for ribulose-bisphosphate carboxylase (temperate.Contig118, 8 ESTs), *psbR* a photosystem II polypeptide (temperate.Contig117, 7 ESTs); *psbP*, a chloroplast precursor gene (temperate.Contig116, 6 ESTs); and a photosystem II thylakoid membrane protein (temperate.Contig115, 6 ESTs).

Different genes of the ribosomal genes category were found in the arctic-enriched and in the temperate-enriched library, but none was especially well covered by ESTs (Table 3.3, Supplementary Tables S3.3 and S3.4). Although the shallow EST coverage suggests that transcript abundance for individual genes was not highly different between the cDNA libraries compared, it is worth noticing that more than a third of the 80 described cytosolic ribosomal proteins (Barakat et al. 2001) were found in one or the other *Oxytropis* enriched cDNA libraries (Table 3.3).

Among the 58 *Arabidopsis thaliana* genes associated to "nucleosome assembly" (GO term 0006334), 10 were in the *Oxytropis* temperate-enriched library. They encode plastidic, mitochondrial and nuclear histone proteins (Table 3.3). Finally, the temperate-enriched library comprised additional genes from a variety of processes, where each gene was not well covered by ESTs.

The results of similarity searches performed between the two libraries indicate that, even with the permissive threshold value in place, only 21 of the 364 temperate unique genes and 19 of the 117 arctic unique genes could be suspected background (potential false positives, Table 3.4), suggesting that the two subtracted libraries are distinct.

The two subtracted libraries comprise numerous unique genes from several gene families. Considering that each subtracted library is composed of two different genomes, the finding of more than two similar genes in one library suggests they can be either orthologous or paralogous genes. Gene families retrieved from the arctic-enriched library include the pathogenesis-related class 10 proteins PR-10 (10 unique genes, Supplementary Table S3.5 for sequence divergence), defensins PDF1 (8 unique genes, Supplementary Table S3.6 for sequence divergence) and KS-dehydrins (11 unique genes, Supplementary Table S3.7 for sequence divergence). Gene families retrieved from the temperate-enriched library belong to chlorophyll

a/b binding proteins (12 unique genes, Supplementary Table S3.8 for sequence divergence), lipid transfer protein LTP (5 unique genes, Supplementary Table S3.9 for sequence divergence), ripening related protein (7 unique genes, Supplementary Table S3.10 for sequence divergence), aluminum induced response ADR6 (3 unique genes, Supplementary Table S3.11 for sequence divergence), specific tissue protein STP (4 unique genes, Supplementary Table S3.12 for sequence divergence), vegetative storage protein-like (3 unique genes, Supplementary Table S3.13 for sequence divergence), and metallothionein Type 1 (3 unique genes, Supplementary Table S3.14 for sequence divergence).

Table 3.3 Indication of the level of differential expression of general gene categories, between arctic and temperate *Oxytropis* species, as suggested by the proportion of Expressed Sequence Tags (ESTs) retrieved from arctic-enriched and temperate-enriched *Oxytropis* plantlets subtractive cDNA libraries [c].

	Arctic		Temperate	
Gene category	Number of genes	EST ratio in enriched library [b]	Number of genes	EST ratio in enriched library [b]
No similarity	31	5.85×10^{-1}	28	5.84×10^{-2}
Unclassified	3	3.55×10^{-3}	79	2.12×10^{-1}
ROS[a] related	4	5.91×10^{-3}	13	2.74×10^{-2}
Histones	1	1.18×10^{-3}	12	2.92×10^{-2}
DNA-proteins interactions	0	0	10	2.00×10^{-2}
Cytosolic ribosomal proteins	3	3.55×10^{-3}	24	5.14×10^{-2}
RNA-proteins and non-cytosolic ribosomal proteins	9	1.65×10^{-2}	14	2.57×10^{-2}
Response to stimulus	56	2.31×10^{-1}	36	1.08×10^{-1}
Secondary	0	0	8	1.46×10^{-2}

metabolism				
Hormones	2	4.73×10^{-3}	16	3.83×10^{-2}
Transport	2	2.36×10^{-4}	13	4.01×10^{-2}
Nucleotides	0	0	1	1.82×10^{-3}
Signaling	1	1.18×10^{-3}	8	1.82×10^{-2}
Protein modifications	1	1.18×10^{-3}	15	2.00×10^{-2}
Nitrogen	1	1.18×10^{-3}	5	2.00×10^{-2}
Lipid	4	7.57×10^{-2}	6	1.46×10^{-2}
Carbohydrates	1	1.18×10^{-3}	18	3.65×10^{-2}
Energy	2	2.36×10^{-3}	62	2.65×10^{-1}
Total	121	846 ESTs	368	545 ESTs

[a] ROS: Reactive Oxygen Species

[b] The ratio of ESTs in a category is expressed relative to the total number of ESTs in the subtracted library of origin, as EST ratio = number of EST sequences in library in this category / total number of EST sequences in library. A total of 846 ESTs compose the final set of unique genes in the arctic-enriched library and 545 in the temperate-enriched library.

[c] EST ratios in enriched libraries resulting from the PCR-select method are not expected to directly reflect transcript abundance in original mRNA population (tester).

Table 3.4 For genes marked as background (potential false positives), the similarity to a gene from the other *Oxytropis* enriched library is presented, in terms of a similarity search E value and in percentage pairwise divergence.

Gene from arctic-enriched library	Gene annotation	General category	Most similar gene from temperate-enriched library [a] (and other similar genes)	E value with TBLASTX [a]	Pairwise divergence % [b]
Oapa1_0329_CZ1835_078.ab1	psaB photosystem i p700 apoprotein a2; electron transport	Photosynthesis	Otpt1_0709_CZ1855_003.ab1	3.00 E-69	0.54
Oapa1_1465_AM23_027.ab1	Carbonic anhydrase 1 (ca1) carbonate dehydratase zinc ion binding	Nitrogen	temperate.Contig24 (and temperate.Contig42)	5.00 E-55	1.90
Oapa1_0946_LS228_065.ab1	Calmodulin; Phosphatidylinositol signaling system	Signaling	Otpt1_1172_CH0101_034.ab1	3.00 E-13	13.98
Oapa1_1518_AM23_079.ab1	Lipid transfer protein precursor	Transport	temperate.Contig102 (and temperate.Contig18, temperate.Contig31, temperate.Contig2, and temperate.Contig106	9.00 E-21	48.18
arctic.Contig46	Pollen coat protein; ABA-inducible protein-like protein	Hormone	temperate.Contig26 (and temperate.Contig40)	7.00 E-67	2.93
Oapa1_1502_AM23_057.ab1	Dormancy/auxin repressed protein	Hormone	Otpt1_1075_CH0101_075.ab1	7.00 E-54	6.11
arctic.contig59	Dehydrin	Stimulus response	temperate.contig22	.	37.7
arctic.Contig34	Pathogenesis related protein; ABA-responsive protein ABR18	Stimulus response	temperate Contig54 (and Otpt1_0104_HFY328_072.ab1)	8.00 E-104	5.24
arctic.Contig39	Pathogenesis related protein; ABA-responsive protein ABR18	Stimulus response	temperate Contig54 (and Otpt1_0104_HFY328_072.ab1)	1.00 E-98	3.23

arctic.Contig9	Pathogenesis related protein; ABA-responsive protein ABR18	Stimulus response	temperate Contig54 (and Otpt1_0104_HFY328_072.ab1)	.	9.42
arctic.Contig61	Pathogenesis related protein; class 10	Stimulus response	temperate Contig54 (and Otpt1_0104_HFY328_072.ab1)	.	3.89
arctic.Contig18	Pathogenesis related protein; class 10	Stimulus response	temperate Contig54 (and Otpt1_0104_HFY328_072.ab1)	3.00 E-87	10.23
Oapa01_0009_CZ1298_033.ab1	Trypsin protein inhibitor 3; Kunitz trypsin protease inhibitor	Stimulus response	Otpt1_0075_HFY328_053.ab1	3.00 E-121	2.01
arctic.Contig58	Acetyl-carboxylase carboxyltransferase beta subunit	Lipid	temperate.Contig41	2.00 E-39	5.83
arctic.Contig57	Acetyl-carboxylase carboxyltransferase beta subunit	Lipid	temperate.Contig41	6.00 E-40	2.88
arctic.Contig6	30S ribosomal protein S12 B	Ribosome and translation	Otpt1_0221_HFY330_051.ab1	4.00 E-05	47.61
arctic.Contig32	30S ribosomal protein s12	Ribosome and translation	Otpt1_0221_HFY330_051.ab1	1.00 E-05	48.62
arctic.Contig49	23S ribosomal RNA; protein orf91; chloroplast genome	Ribosome and translation	Otpt1_1376_CH0101_033.ab1	3.00 E-80	1.42
Oapa1_0163_HFY329_045.ab1	Weakly similar to a Histone deacetylase and COR8.6 protein	Histone	Otpt1_0616_CZ1855_037.ab1	3.00 E-23	9.76
arctic.Contig29	Phospholipid hydroperoxide glutathione peroxidase	ROS	Otpt1_0294_CZ1835_003.ab1 (and Otpt1_1358_CH0101_018.ab1)	4.00 E-65	32.83
Oapa1_0354_CZ1835_094.ab1	Rhicadhesin receptor precursor (Germin-like protein); Cupin domain	ROS	temperate.Contig64 (and Otpt1_1500_AM21_039.ab1)	2.00 E-31	3.02

arctic.Contig48	NA	Unknown	Otpt1_0918_CH0101_064.ab1	-	40.97

[a] Genes were marked as potential background (false positive) based on BLASTN or TBLASTX permissive E value threshold below 1 e-04, however, only the TBLASTX E value is reported in the table for brevity.

[b] Percentage pairwise divergence was calculated with uncorrected p-distance, with gap pairwise deletion in MEGA.

3.4.2 Real-time RT-PCR of selected genes supports differential gene expression

Transcript levels of four genes were further investigated using real-time RT-PCR for all four species (*O. arctobia, O. campestris* subsp. *johannensis, O. maydelliana* and *O. splendens)*, each grown in arctic and in temperate simulated conditions (Fig. 1). By sequencing, target genes were confirmed to exist in all four genomes, and primers were confirmed to amplify a single product (Supplementary Table S3.2, GenBank accession numbers HM107135 to HM107155).

For all four genes, transcript abundance as estimated by real-time RT-PCR (Fig. 3.1) appears higher in the combined two cDNAs that were used in the tester library (Table 3.1), than abundance in the driver library. This observation supports the results found with PCR-select method.

Real-time RT-PCR data also suggest that each two arctic species maintain a similar level of KS-dehydrin arctic.contig.47 transcript abundance under the arctic (Fig. 3.1a) and the temperate (Fig. 3.1b) conditions, whereas for the temperate species, abundance appears more elevated in arctic conditions (3.7 or 3.9 fold higher than actin, Fig. 3.1a) than in temperate conditions (0.5 or 1.4 that of actin, Fig. 3.1b). This KS-dehydrin may therefore be constitutive in arctic species, but cold induced in temperate species.

Arctic.contig61 and arctic.contig13/36 are two gene copies of the PR-10 family (pathogen-related proteins, class 10). Both genes are present in the genome of all four *Oxytropis* species (Supplementary Table S3.2) but the real-time RT-PCR data (Fig. 3.1c, d, e, f) reveal they have different expression patterns. In arctic conditions, abundance of arctic.contig61 gene is high in the arctic *O. arctobia* (323 fold higher than actin), and is also noticeable in the two temperate species, but to a

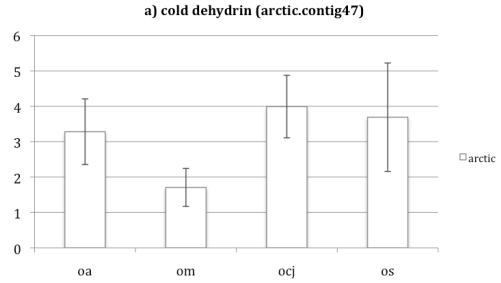
lesser extent (15.6 and 3.2 fold higher than actin, Fig. 3.1c). In the temperate conditions, paralleling the situation observed in the arctic condition, the arctic.contig61 gene is only abundant in *O. arctobia* (57.8 fold higher than actin, Fig 3.1d), but apparently less than in the arctic conditions.

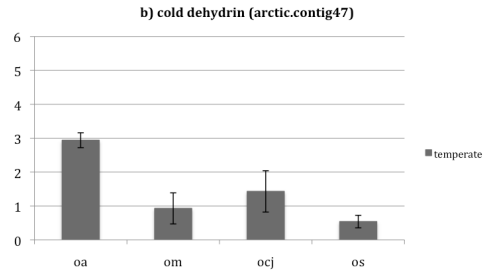
The other PR-10 gene surveyed was called arctic.contig13/36 because, according to gDNA fragments (Archambault and Strömvik 2012b), it corresponds to the 5' end of arctic.contig13, but to the 3' end of arctic.contig36. In arctic conditions (Fig 3.1e), abundance of this transcript is low (between 0.2 and 0.6 that of actin) in three species and is null in *O. arctobia*. In the temperate conditions, arctic.contig13/36 is absent from all transcriptomes (Fig 3.1f). The real-time RT-PCR results suggest there is a contrasting expression pattern between the two arctic species for these two very similar PR-10 genes (8% to 15% sequence divergence including the intron region, Supplementary Table S3.15): *O. arctobia* has transcripts only from the arctic.contig61 PR-10 at a high level, while *O. maydelliana* has transcripts from both PR-10, but at low abundance.

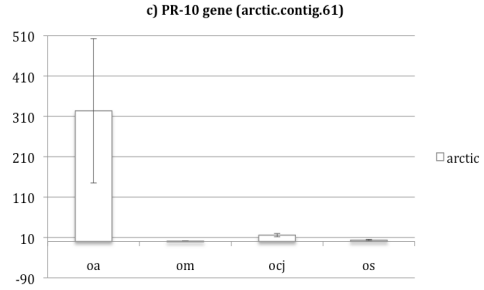
The temperate.contig119 represents a chloroplast photosystem II light harvesting complex protein type I (*lhcbI*) gene. In temperate conditions, transcript abundance appears higher in *O. splendens* (70 fold higher than actin) than in the other three species (Fig 3.1h). Abundance in the arctic condition is, however, moderate (9.7 to 21 fold higher than actin) in all four species (Fig 3.1g). The real-time RT-PCR data for these four genes supports results found with the PCR-select method. It shows, however, that gene expression can be species-specific, rather than common in the two species from the same environment.

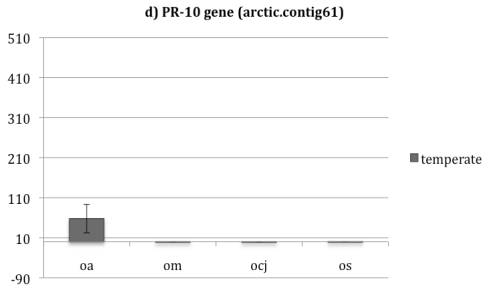
Figure 3.1 Transcript abundance of selected genes compared in transcriptomes of two arctic *Oxytropis* species (*O. arctobia* and *O. maydelliana*) and two temperate species (*O. campestris* subsp. *johannensis* and *O. splendens*) under two climatic conditions.

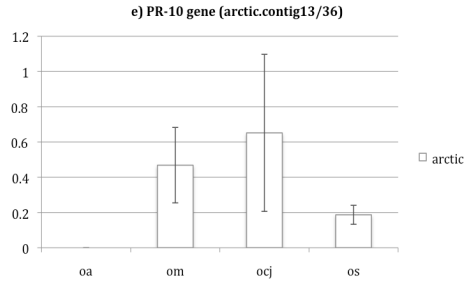
Real-time RT-PCR was used to measure gene expression in cDNA. Values on the Y-axis are the mean ratio of the selected gene relative to actin expression. Actin was used as a normalizing gene ±1 standard deviation of three biological replicates and two technical replicates. Transcript levels of arctic.contig47 (KS-dehydrin) under a) arctic and b) temperate conditions; arctic.contig61 (PR-10) under c) arctic and d) temperate conditions; arctic.contig13/36 (PR-10) under e) arctic and f) temperate conditions; temperate.contig98 (light harvesting protein I, lhcbI) under g) arctic and h) temperate conditions. Abbreviations: oa: *O. arctobia*; om: *O. maydelliana*; ocj: *O. campestris* subsp. *johannensis*; os: *O. splendens*. The white bars represent transcript levels of plantlets grown in simulated arctic conditions, and the dark grey bars, the levels in plantlets grown in simulated temperate conditions.

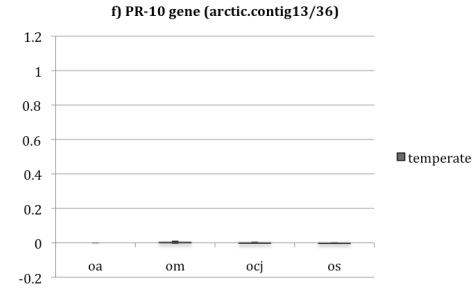

a) cold dehydrin (arctic.contig47)
6
5
4
3
2
1
0
oa
om
ocj
os
arctic

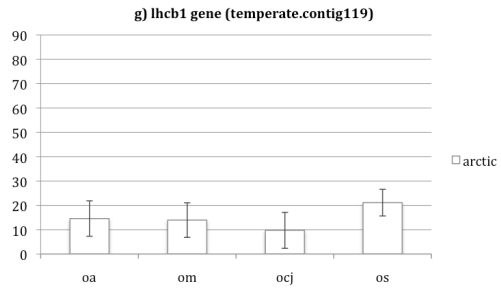

b) cold dehydrin (arctic.contig47)
6
5
4
3
2
1
0
oa
om
ocj
os
temperate

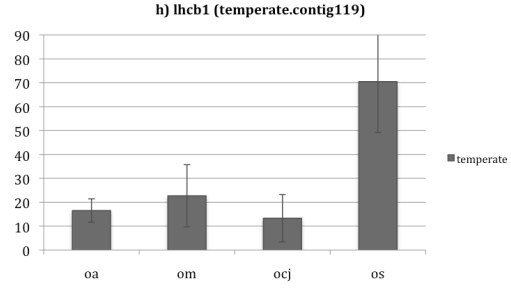

c) PR-10 gene (arctic.contig.61)
510
410
310
210
110
10
-90
oa
om
ocj
os
arctic


d) PR-10 gene (arctic.contig61)
510
410
310
210
110
10
-90
oa
om
ocj
os
temperate


e) PR-10 gene (arctic.contig13/36)
1.2
1
0.8
0.6
0.4
0.2
0
oa
om
ocj
os
arctic


f) PR-10 gene (arctic.contig13/36)
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
oa
om
ocj
os
temperate


g) lhcb1 gene (temperate.contig119)
90
80
70
60
50
40
30
20
10
0
oa
om
ocj
os
arctic


h) lhcb1 (temperate.contig119)
90
80
70
60
50
40
30
20
10
0
oa
om
ocj
os
temperate

Library subtraction (PCR-select) and real-time RT-PCR methods do not evaluate the same expression data and results are not expected to be strictly comparable. A library subtraction enrichment (PCR-select) should roughly reflect difference in transcript abundance between two compared libraries (tester and driver, Table 3.1), but is not necessarily expected to reflect abundance in the tester library. Transcript abundance in a non-manipulated cDNA library may be better estimated by real-time RT-PCR. Here, the *Oxytropis* data from the two methods were visually compared for similarities in the suggested trends in transcript abundance. Among the four genes surveyed with real-time RT-PCR, the arctic.contig61 gene was the most abundant with 323 fold more transcripts than actin (Fig. 3.1) in the arctic cDNAs that were used as tester for the PCR-select. That gene also showed the greatest abundance difference between the two transcriptomes that were used as tester and as driver, being 323 and 0.2 that of actin (Fig. 3.1), respectively. In addition, that gene had the highest ESTs ratio from the PCR-select constituting more than 10% of the EST collection (94/846 ESTs, Supplementary Table S3.3). Similarly, the arctic.contig13/36 is, among the four genes surveyed, the least abundant in the cDNAs used as tester (Fig. 3.1e), and also shows low abundance difference between the two transcriptome used as tester and as driver, being 0.46 and 0.01 that of actin (Fig 3.1), respectively. In addition, that gene arctic.contig13/36 had a low ESTs coverage from the PCR-select, with only 3 (Supplementary Table S3.3). Overall, these comparisons suggest a tendency for genes highly covered by ESTs in a PCR-select enriched library (Supplementary Table S3.3 and S3.4) to also show a high difference in transcript abundance between original mRNA populations (real-time RT-PCR data, Fig. 3.1).

3.5 Discussion

3.5.1 Arctic *Oxytropis* plantlets exhibit a lower expression for photosynthesis related genes, typical of cold acclimation

In order to shed light on potential molecular adaptations that arctic plants have developed, we have sequenced and compared the subtracted plantlet transcriptomes of arctic and temperate *Oxytropis* legume species. The subtraction experiment is supported by real-time RT-PCR, for a subset of genes.

These plants express distinct transcriptome signatures in their natural environment that show both typical and novel features compared to latitudinal or altitudinal gradients in transcriptome variation in other plant lineages (Swindell et al. 2007; Holliday et al. 2008; Voelckel et al. 2008), or to plant species adapted to other abiotic stresses (Taji et al. 2004; Brosche et al. 2005; Filatov et al. 2006; Hammond et al. 2006; Knight et al. 2006; Lai et al. 2006; van de Mortel et al. 2006). Given that the important constraints to plant growth in the arctic are low summer temperature and frequent summer frosts (Savile 1972), arctic species are expected to show adequate expression of cold treatment genes. Gene expression reorganization following cold stress is now well characterized in model (Hannah et al. 2005) and agronomical (Cheng et al. 2007) plant species, and many of the differentially expressed genes between arctic and temperate *Oxytropis* species conform to this pattern. In *Arabidopis*, genes upregulated after long term cold exposure are mainly of the stress response category, especially the hydrophilic COR/LEA proteins (Hannah et al. 2005), and some are found in the *Oxytropis* arctic-enriched library.

Several genes downregulated after long term cold exposure in *Arabidopis*, such as photosynthesis related genes, carbohydrate metabolism, GDSL-motif lipase,

hormone metabolism and oxidative regulation genes (Hannah et al. 2005) are found in the *Oxytropis* temperate-enriched library. For one species, the temperate *O. splendens*, the transcript level of the light harvesting gene *lhcbI* (calculated by real-time RT-PCR) appears higher in the temperate condition than in the arctic condition. Although it is mostly species-specific, this finding is consistent with the deterioration of photosynthetic capacity following cold exposure in most plants (Savitch et al. 2001; Stitt and Hurry 2002; Walters 2005) and with a plastic response of arctic plants that modify their optimum temperature for photosynthesis and carbon integration within a few days after being placed in warmer growth conditions (Pyankov and Vaskovskii 1994).

3.5.2 Defence response is a prominent characteristic of the arctic plantlet transcriptome

The constraints on plant growth imposed by the arctic environment extend beyond cold temperatures, and also include very short growing season, strong winds, low light intensity but long days. It is therefore expected to see a set of expressed genes in the arctic *Oxytropis* plantlets unique and different from a cold treated temperate model plant. The most striking feature of the *Oxytropis* arctic-enriched library is its enrichment in novel genes, and in genes of the "response to stimulus" category, indicating that, compared to temperate species, arctic *Oxytropis* under arctic simulated growth conditions expressed more of two unknown genes, one uncharacterized gene, defensin (PDF1), pathogenesis-related proteins (PR-10), KS-dehydrins, early light inducible (ELIP) genes. These four response to stimulus genes are not, as a group, typical of those upregulated during cold acclimation in temperate plants.

In other species, stress response genes show a substantial among taxa variation in gene expression (e.g. Chen et al. 2005), even under non-stress conditions. Genes of this category also exhibit, in favourable conditions, an elevated expression for populations or species adapted to adverse environmental conditions (Taji et al. 2004; Beritognolo et al. 2008), whether for drought (Brosche et al. 2005; Knight et al. 2006), metal (van de Mortel et al. 2006), saline (Taji et al. 2004) or sub-arctic growth condition (Holliday et al. 2008).

We show that the PR-10 gene family is, after two unknown genes, one of the largest overrepresented in arctic-enriched *Oxytropis* transcriptomes*;* but that expression of the different copies is species-specific. Notably, the two arctic species do not show a common pattern. The *O. arctobia* plantlet has transcripts for a PR-10 paralog different than in other three species, which could indicate that subfunctionalization has been effective during the course of *Oxytropis* genome evolution, as described for groups of species comprising polyploids (Adams 2007). The original biological roles attributed to PR-10 proteins were in plant defence with antibacterial, antimicrobial (Pinto and Ricardo 1995; Broderick et al. 1997; Pinto et al. 2005) and ribonucleolytic functions (Bantignies et al. 2000), but recent evidence suggests additional functions in cold response (Goulas et al. 2007), development (Iturriaga et al. 1994; Sikorski et al. 1999; Bantignies et al. 2000; Pinto et al. 2005), hormone binding (Fernandes et al. 2008) and secondary metabolism (Liu and Ekramoddoullah 2006). Among all these functions, the ones related to pathogen response are less likely to have played a role in adaptation to the Arctic given that potential pathogens (Strathdee and Bale 1998) and fungal symbionts (Kytoviita 2005) have low diversity and abundance in that area.

Plant defensins are other “response to stimulus” genes overrepresented in the arctic *Oxytropis* plantlets transcriptome. Similarly to the PR-10, plant defensins (reviewed in Thomma et al. 2002) were first described as antifungal proteins (Terras

1995) but were also shown to be developmentally regulated in healthy legumes (Ishibashi et al. 1990; Hanks et al. 2005). Plant defensins show differential expression in many pairs of closely related species (Hanks et al. 2005; Hammond et al. 2006; van de Mortel et al. 2006; Holliday et al. 2008), to the favour of an increased expression in the stress adapted species or population. Defensins are abundant at seed germination and could protect from soil born pathogens (Hanks et al. 2005; Carvalho and Gomes 2009), which is compatible with the expression detected in very young *Oxytropis* plantlets. However, as for the PR-10, other biological roles, such as zinc tolerance (van de Mortel et al. 2006), regulation of the ascorbic acid redox state and even self-incompatibility (reviewed in Carvalho and Gomes 2009), may also explain the differential expression between *Oxytropis* species.

KS-dehydrins is another "response to stimulus" gene family largely present in the arctic *Oxytropis* transcriptome. These genes encode group 2 (D-11) LEA proteins, which show high sequence divergence even between closely related species (Battaglia et al. 2008). The most similar described genes are from legume species (the *Glycine max* SLTI629 ABQ81887.1 and SRC1 BAA19768.1; the *Medicago truncatula* cas15 ABX80065.1; the *Medicago sativa* BudCAR5 AAF33785.1 and CAR1 AAC25776.1), and from the Brassicaceae model plant *Arabidopsis thaliana* as well (At1g54410, a dehydrin family protein). A certain level of constitutive expression for dehydrins was described (Boudet et al. 2006), in addition to its induction by cold, heat, drought, wounding and virus infection in soybean (Takahashi and Shimosaka 1997) and in *Medicago* (Chen et al. 2008a; Pennycooke et al. 2008). Our finding that expression of a KS-dehydrin gene may be less responsive to the growth conditions in the arctic species than it is in the temperate species is in agreement with reports on a less responsive expression of "response to stimulus" genes in stress adapted plant species (Taji et al. 2004; Brosche et al. 2005).

The proposed role of dehydrins in drought tolerance (Close 1996) is desirable in an environment where water is limited (Aiken et al. 2007) and where frequent summer frosts can induce ice formation in the apoplast leading to cellular dehydration.

A concurrent overexpression of the gene families PR-10, defensin and KS-dehydrin, is a novel feature for plants adapted to adverse environmental conditions. Other "response to stimulus" genes are overrepresented in arctic *Oxytropis*, although with a less striking expression difference, and several of these were also overexpressed in other stress tolerant species. The early light inducible proteins (ELIP) are overrepresented in arctic *Oxytropis*, a pattern described earlier (Lai et al. 2006), but that is not universal since they were also overexpressed in the lower altitude *Pachycladon fastigata* (Voelckel et al. 2008) or the Californian *Picea* (Holliday et al. 2008).

3.5.3 Ribosome biogenesis and assembly genes are differentially expressed between arctic and temperate species

The library subtraction identified that a third of the described cytosolic ribosomal genes (Barakat et al. 2001) are differentially expressed between arctic and temperate *Oxytropis*. Many genes are more abundant in the temperate plantlets transcriptome, but a few are on the contrary more abundant in the arctic plantlets. The level of differential expression is, however, likely low because individual genes are covered by only a few ESTs. Nonetheless, if proved valid, this is a novel and not previously described pattern for arctic plants, which would suggest that organization of the cytosolic ribosome protein complex might play a role in long term adaptation of plants to the arctic, or in specific response to arctic conditions. In other species, it has been shown that several of these genes either increase (Saez-Vasquez et al.

2000; Kim et al. 2004) or decrease in (Berberich et al. 2000; Swindell et al. 2007) expression following stresses; or are differentially regulated during development (McIntosh and Bonham-Smith 2006; Whittle and Krochko 2009). Furthermore, differential expression of ribosomal genes among other population and species has been detected (Filatov et al. 2006; Hammond et al. 2006; Holliday et al. 2008; Voelckel et al. 2008), where, as in *Oxytropis,* more ribosomal genes are underexpressed than overexpressed in the species adapted to more stressful environment. This observation could reflect the observed slower growth of arctic plants (Bliss and Gold 1999; Aiken et al. 2007) in their natural habitat.

3.5.4 Nucleosome assembly genes are underrepresented in the arctic transcriptome

Histone genes are likely differentially expressed, since 10 out of the 58 that exist in the well-studied *Arabidopsis thaliana* are represented in the temperate-enriched library. This suggests that relative lower transcripts abundance for genes of this biological process in the arctic *Oxytropis* plantlets may participate in their distinctiveness. Although a close look at differentially expressed genes in other plant species adapted to stressful conditions reveals anecdotic overexpression of some histone genes and underexpression of others (Filatov et al. 2006; Hammond et al. 2006; Holliday et al. 2008), we have not found examples where this category of genes is an important feature of a specialized transcriptome. In addition to the differential expression we show here in *Oxytropis* for histone genes, previous findings describing that different histone genes can be differentially regulated during plant development (Huh et al. 1995) or stress response (Kapros et al. 1992) suggest that regulation of expression for genes related to chromatin assembly may also participate in plant

adaptation to environmental conditions, rather than be a simple cause of slower cell division for arctic species (Meshi et al. 2000).

3.5.5 Conclusion

This is the first report to our knowledge on gene expression profiles of an arctic plant and our findings are supported by previous reports on plant adaptation to stressful environmental conditions. Arctic *Oxytropis* species, as opposed to temperate ones, overexpress novel genes, stimulus response genes such as certain PR-10 genes, defensin and KS-dehydrins and under express photosynthesis and histone genes. Real-time RT-PCR results also show that a cold dehydrin may have participated in adaptation to the Arctic because it is constitutive in the arctic species, and cold induced in the temperate species.

3.6 Acknowledgements

This work was supported by a grant from Natural Sciences and Engineering Research Council of Canada (NSERC) to MVS. AA was supported by Le Fonds québécois de la recherche sur la nature et les technologies (FQRNT), Eben-Hopson and Girardin-Vaillancourt scholarships. Field trip was made possible by Northern Scientific Training (NSTP) grants from the Department of Indian and Northern Affairs Canada, and by Centre Sève (FQRNT). We would like to thank Dr. Susan Aiken, Dr. Danielle Prévost, Dr. Carolyn Parker, Dr. Tracy Sterling and Plant Gene Resources of Canada (Agriculture and Agri-Food Canada, Saskatoon Research Centre) for providing *Oxytropis* seeds and Christine Ide for technical help.

Chapter 4

4 Connecting text to Chapter 4

In the comparative transcriptomic study between arctic and temperate *Oxytropis* plantlets presented in Chapter 3, several strongly differentially expressed genes, members of gene families, were identified. Interpreting these sequence data was, however, limited because they were generated from multi-species cDNA pools. Putative genes could consequently not be assigned to a single species and missed the intron region. Chapter 4 is based on sequences isolated from gDNA, and presents phylogenetic, and codon selection analyses three of these gene families: dehydrins and pathogenesis related class 10 (PR-10) that were expressed in the arctic species, and ripening-related proteins that were expressed in the temperate species.

The article "The Y-segment of novel cold dehydrin genes is conserved and codons in the PR-10 genes are under positive selection in *Oxytropis* (Fabaceae) from contrasting climates" was published in Molecular Genetics and Genomics online first December 19, 2011 (DOI: 10.1007/s00438-011-0664-6). It is covered by copyrights, and is included entirely in Supplementary Figure S4.2, with kind permission from the publisher Springer under license number 2835461220405 provided by the Copyright Clearance Center. Nucleotide sequence data reported are available in the DDBJ/EMBL/GenBank databases under the accession numbers HQ731797 to HQ731908. An augmented version is presented in Chapter 4, and includes data on percentage pairwise sequence divergence that corroborate the original publication.

4.1 The Y-segment of novel cold dehydrin genes is conserved and codons in the PR-10 genes are under positive selection in *Oxytropis* (Fabaceae) from contrasting climates

4.1.1 Abstract

While the arctic flora is particularly threatened by climate changes, the molecular aspects allowing colonization of this harsh environment remain largely enigmatic. Genes with a likely functional or evolutive role for arctic *Oxytropis* (Fabaceae) were previously discovered given a sharp differential expression between arctic and temperate species. Here, we analyze gene duplication patterns, and positive and negative selection between genes from species of contrasting environments, which can reveal potential gene functions. Genes were amplified and sequenced from two arctic (*O. arctobia* and *O. maydelliana*) and two temperate (*O. campestris* subsp. *johannensis* and *O. splendens*) species. Detection of codons under positive or negative selection and phylogenetic analyses were used to further elucidate pathogenesis-related class 10 (PR-10), ripening-related proteins, dehydrin gene families and light-harvesting complex (lhcaIII and lhcbI) genes from *Oxytropis*. Overall, results showed that the three gene families duplicated prior to the *Oxytropis* genus diversification; that genes overexpressed in arctic species evolve under higher constraints at the sequence level in these species; that evolving novel protein variants in PR-10 genes characterizes initial adaptation to the Arctic, and that *Oxytropis* dehydrins are of a novel (K-like - Y_4 – K– S) structure, where the Y-segment is under stringent evolutive constraints in the arctic species. This suggests a scenario not previously described for arctic plants, where gene duplications precede arctic species establishment, and where genes later become both highly expressed and under stringent constraints in the arctic species.

4.2 Introduction

The understanding of the molecular mechanisms behind plant adaptation to the Arctic is still fragmentary. The already species poor arctic flora (Callaghan et al. 2004) is further particularly threatened by climate changes (Totland and Alatalo 2002; Marchand et al. 2006; Walker et al. 2006). Nevertheless, arctic plants serve important purposes, such as feeding mammalian herbivores and insect pollinators. Using library subtraction and clone sequencing (ESTs), we recently reported that arctic plantlets preferentially express response to stimulus genes (especially PR-10, dehydrins and defensin gene families) (Archambault and Strömvik 2011). On the other hand, temperate plantlets express photosynthesis genes (Archambault and Strömvik 2011), including *lhcbI* and *lhcaIII*, that encode proteins of the light-harvesting complex, which function as a light receptor to capture and deliver excitation energy to photosystems (Swarbreck et al. 2008). While the photosynthesis genes' differential expression may be driven by temperature (Hannah et al. 2005) and light conditions, a biotic reason for the high abundance of response to stimulus genes in arctic plants is not obvious because the Arctic is not rich in fungal symbionts, endophytes and pathogens (Kytoviita 2005; Higgins et al. 2007). Since other genomic evolutive features, such as accelerated sequence evolution (Kliebenstein et al. 2006) or gene duplications (Flagel and Wendel 2009) commonly accompany important differential gene expression, these will be explored in the present work, to give insight into evolution of the PR-10, and KS-dehydrin genes, overrepresented in arctic *Oxytropis* transcriptomes, and into the ripening-related proteins, *lhcbI* and *lhcaIII* genes, overrepresented in temperate *Oxytropis* transcriptomes.

The *Oxytropis* genus is predominantly distributed in the temperate and boreal regions of North America and Northern Asia, and also includes 44 arctic

species (Elven 2007). The genus comprises many polyploid species, but for almost all, it is not know whether they are of allopolyploid or autopolyploid origin. To our knowledge, this information is only known in a rare Russian *Oxytropis* species, found to be autotetraploid (Kholina et al. 2004).

Detection of positive and negative selection in gene sequences can reveal how selection has shaped protein-coding sequences. Computational tools for this are applicable to a wide array of questions, because they require only a sequence alignment and no assumptions about population demography (Kosakovsky Pond et al. 2005; Nielsen 2005). They were shown to be valid and insightful in various contexts and from various genomes, from plants (Mondragon-Palomino et al. 2009; Zamora et al. 2009) to fishes (van der Aa et al. 2009; Elmer et al. 2010) and fungal pathogens (Aguileta et al. 2010).

Many PR-10 (pathogenesis-related class 10) and ripening-related genes were identified within the arctic and temperate transcriptomes, respectively (Archambault and Strömvik 2011). These stimulus response genes are far better known for their allergenic effect in human (Berkner et al. 2009), while their functions in plants remain somewhat elusive. KS-dehydrins are other response stimulus genes overrepresented in arctic *Oxytropis* plantlet transcriptomes (Archambault and Strömvik 2011). Dehydrin proteins are small hydrophilic, heat-stable proteins known as LEA D-11 family (reviewed in Battaglia et al. 2008). They are expressed following dehydration, salinity and low temperature, but are also present under optimal growth condition. Dehydrin proteins are found in most plant genomes, but vary widely in amino acid sequences between species; and they are characterized by the presence of repeats such as the lysine rich K-segment (reviewed in Battaglia et al. 2008).

The differential expression of multiple copies of many different stimulus response genes families was an important finding from our initial comparative

transcriptome study (Archambault and Strömvik 2011), but was gained from transcriptome data. The general objective of the present work is to characterize key evolutionary features, from gene sequences isolated from gDNA, in arctic and temperate *Oxytropis* species. We achieve this goal by detection of codons under selection as well as phylogenetic reconstructions, for the PR-10, ripening-related proteins, KS-dehydrins gene families, and for the photosynthesis-associated proteins LHCAIII and LHCBI. We aimed at elucidating specific questions: (1) Is there evidence for more than one copy in the genome of the four *Oxytropis* surveyed for the three gene families surveyed, (2) Do PR-10 and KS-dehydrin genes, found overexpressed in arctic species, have selective pressure different in arctic species, compared to temperate species? and (3) Do PR-10 and ripening-related proteins, two related gene families that show opposite expression profile in *Oxytropis*, have similar or contrasting codon selection patterns?

The results point to molecular characteristics of arctic plants, such as initial duplications in response to stimulus genes, and subsequent increase of positive selection in only some of the gene families, that may be involved in their exceptional resistance to harsh climatic conditions. This study provides evidence for molecular characteristics that could help identify the possible survival capacity of the different arctic plant species in a context of climate changes.

4.3 Materials and methods

4.3.1 Gene isolation from genomic DNA

4.3.1.1 Plant genomic DNA material

Oxytropis seeds were collected by the authors, kindly provided by collaborators or ordered from plant nurseries. The seed sources and ploidy levels are listed in Supplementary Table S4.1. Seeds of the arctic species *O. arctobia* (diploid) and *O. maydelliana* (dodecaploid); the temperate species *O. campestris* subsp. *johannensis* (hexaploid) and *O. splendens* (diploid); and the nearly subtropical *O. lambertii* Pursh (hexaploid) were scarified, sterilized and stratified at 4°C on 1/2 MS Basal Medium (Sigma, Oakville, Ontario) agar plates. Plates were transferred to growth chambers (16 h of light, (225 μmol/m^2/s) at 22˚C, and 8 h darkness at 18˚C). Plantlets were collected when they developed two leaves, photographed and immediately placed in liquid nitrogen. Total DNA was extracted with the DNeasy plant mini kit (catalog number 69104 Qiagen, Mississauga, Canada).

4.3.1.2 Primer design

Primers were made to be specific for genes represented by contigs, which were assembled from *Oxytropis* plantlet Expressed Sequence Tags (ESTs) (Archambault and Strömvik 2011) but were designed to allow consistent amplification from the four *Oxytropis* species genomic DNA. Targeted contigs correspond to PR-10 genes (arctic.contig.13, arctic.contig.18, arctic.contig.34, arctic.contig.36 and arctic.contig.61); to ripening-related protein genes (temperate.contig.93, and temperate.contig.101); to KS-dehydrin genes (arctic.contig.47, arctic.contig.59, arctic.contig.27 and arctic.contig.25); to light-harvesting *lhcbI* gene (temperate.contig.119); and to the light-harvesting complex a (*lhcaIII*) gene (temperate.contig.110). Twenty two primers were designed for PR-10 genes, twelve for ripening-related proteins, and fifteen for KS-dehydrins, using Primer3 (Rozen and Skaletsky 2000; Koressaar and Remm 2007) implemented in

the Geneious software (Drummond et al. 2008), and were generally 22 to 24 bp (base pair) long (Supplementary Table S4.2).

4.3.1.3 PCR amplification of selected genes from genomic DNA

The Amplitaq Gold (catalog number 4311816 Applied Biosystems by Life Technologies, Carlsbad, California) was the DNA polymerase generally used, except with some primer pairs (lhcaIII_110_1F with lhcaIII_110_2R; lhcbI_119_1F with lhcbI_119_2R; arct_cold27_1F with arct_cold25_2R; arct_cold59_3F with arct_cold59_4R and PR10_13_1F with PR10_13_2R) that needed a higher annealing temperature to improve specificity and where the IProof polymerase (catalog number 172-5301, Bio-Rad laboratories Canada, Mississauga, Ontario) was used. The IProof polymerase amplified stronger fragments with low unspecific product at annealing temperatures between 68˚C to 71˚C, whereas this temperature was 59˚C to 66˚C for the Amplitaq Gold (Supplementary Table S4.2). Amplifications from gDNA of each *Oxytropis* species were performed in a 50 μl volume. Reactions using the IProof polymerase were made at a final concentration of 2.6 mM $MgCl_2$, 0.2 mM dNTP and 0.4 μM primers and used 0.8 units of enzyme and 30 ng DNA. The thermocycling program applied was: 98˚C for 30s; followed by 30 to 32 cycles of denaturation at 98˚C for 6s; annealing (temperature mainly 69˚C or 71˚C) for 30 s; extension at 72˚C for 2 min; and a final extension of 7 min at 72˚C. Amplification reactions using the Amplitaq Gold were made at a final concentration of 1.5 mM $MgCl_2$, 0.2 mM dNTP 0.4 μM primers, 1.25 units of enzyme and 30 ng DNA (or less when quantity was limited). The thermocycling program applied began by 95˚C for 5 min; followed by 30 to 32 cycles of denaturation at 98˚C for 2 min; annealing (temperature mainly 59˚C, 63˚C or

66˚C) for 30 s; extension at 72˚C for 3 min; ended by final extension of 7 min at 72˚C.

In some cases, a PCR reaction resulted in two amplified fragments. When they could be clearly separated by gel electrophoresis, the different amplified fragments were cut from the gel, purified using QIAquick (cat no 28706 Qiagen, Mississauga, Canada), and sent for sequencing. A nested amplification approach was applied when amplification resulted in the absent, faint or multiple bands. A first round of amplification was performed with primers specific for a larger fragment, using iProof polymerase in a 10 μl reaction for only 20 cycles. These amplified products were then used as template for a second round of amplification using internal primers, specific for a shorter fragment, using Amplitaq Gold polymerase in a regular 50 μl reaction. All PCR reactions that resulted in a single band, as determined by gel electrophoreses, were sent unpurified, for sequencing from both ends (Sanger sequencing on a 3730xl DNA Analyzer of Applied Biosystems at the McGill University and Génome Québec Innovation Center).

4.3.2 Sequence assembly and editing

In order to reconstruct, for a given genome, the sequence of each gene copy of a gene family while ensuring that a gene sequence is not present more than once in the subsequent analyses, sequence assembly was undertaken in two steps, as follows. In the first step, the forward and reverse chromatogram from each PCR reaction were assembled in Geneious (Drummond et al. 2008) with default values and the sequences was edited by eye. Then, in the second step, sequences from different PCR reactions were collapsed if they likely correspond to the same gene. This was determined by building an initial neighbor-joining phylogenetic tree in Geneious with all the primary assemblies. Secondary assembly was only performed

on near-identical sequences amplified from a same plantlet that were very closely positioned in the initial neighor-joining tree. Sets of sequences that filled these criteria were subsequently treated as a single gene. This secondary assembly step was conservative, and we never assembled sequences with genuine differences. As an example, primer pairs arct_PR10_61_1F with arct_PR10_61_16R and arct_PR10_61_11F with arct_PR10_61_4R amplify the same genomic fragment from *O. splendens* plantlet number 44g, but amplify two distinct fragments from *O. maydelliana* plantlet number 40e.

4.3.2.1 Outgroup sequences identification, coding sequence localization, and protein structure estimation from related sequences

Outgroups for each gene and gene family were identified in 2010 with tblastx (Blosum 62 matrix, *E* value threshold 0.1) similarity search in NCBI public databases, limited to the Fabaceae sequences (Sayers et al. 2010), and in Phytozome v5.0 online database for *Medicago truncatula* and *Glycine max* genomes (Schmutz et al. 2010). Legume gene families already grouped and available for these two genomes in Phytozome database were downloaded. Ripening-related sequence from *Arabidopsis* (Brassicaceae) At1g70830 (*MLP28*) was also included in the alignment and initial phylogenetic reconstructions because it is well characterized (Lytle et al. 2009). Sequences were first aligned using the Geneious alignment algorithm with a cost matrix of 65% similarity and the alignment was manually refined when necessary, considering the *in silico* translated amino acid sequences. Intron position for all *Oxytropis* genomic sequences isolated was located by comparing to the similar publicly available annotated genes. The full gene sequence (including 5'

UTR, 3' UTR and intron) was considered for recombinant detection and pairwise percentage divergence calculation, but only the coding sequence was used for detection of sites under selection and phylogenetic analyses.

The protein data bank (http://www.pdb.org/pdb/home/home.do (Berman et al. 2000) database was searched (in July 2010) to compare *Oxytropis* sequences to similar proteins with experimentally determined structures, using blast similarity search (Altschul et al. 1997). Protein structures as well as positively or negatively selected codons were mapped on the aligned coding sequences.

4.3.3 Detection of recombination

Recombinant sequences may be expected in the present datasets, which includes sequences isolated by PCR amplification from polyploid genomes of genes members of gene families. Recombinants may reflect naturally occurring rare events that took place *in planta,* or may be induced during the PCR amplification step, and hence represent chimeric sequences. Although the processes generating recombinant sequences are not distinguished from one another in the present case, such sequences may complicate downstream sequence analyses. Here recombination was detected in a two-step process. First, recombination in the sets of aligned sequences was detected, and the breakpoint was inferred with the GARD (Genetic Algorithm Recombination Detection) algorithm of the HyPhy package (Kosakovsky Pond et al. 2006). Second, to identify sequences that may be chimeras produced during the PCR amplification, each set of three or four sequences from a single plantlet were analyzed with the RDP tool (Martin et al. 2010).

4.3.4 Detection of sites under positive and negative selection

The three approaches used to detect codons under negative or positive selection, were the single-likelihood ancestor counting, (SLAC), fixed effects likelihood (FEL) and random effects likelihood (REL), all implemented in the online version (http://www.datamonkey.org) of the HyPhy package (Kosakovsky Pond and Frost 2005a). Selection is detected at codons where nonsynonymous mutations significantly differ from synonymous ones. These algorithms were used because they are conservative (Cavatorta et al. 2008; Zamora et al. 2009). Since recombination can induce false positives (Kosakovsky Pond et al. 2006) in codon selection detection, chimeric sequence (identified with RDP tool) were excluded. The alignment was also divided at the recombination break point when recombination was still detected (with GARD tool), and partitions were analyzed separately for codon selection. In those cases, our approach was conservative, and we considered only the codons under selection in the partitions and in the full-length analyses. *Oxytropis* sequences were analyzed altogether, as well as temperate and arctic sequences separately. The *Lupinus* and *Pisum* PR-10 and the *Medicago* ripening-related sequences were analyzed separately as well. In KS-dehydrins, many sequences were truncated upstream position 27, and these sites were not considered.

4.3.5 Phylogenetic analyses

Phylogenetic analyses were performed on the coding region only (excluding intron, 5' and 3' UTR where alignment was not reliable) using the PHYML algorithm (Guindon and Gascuel 2003) implemented in Geneious, in order to clarify the relative timing and pattern of duplication of *Oxytropis* genes families. Bayesian posterior probabilities were estimated using MrBayes 2.0 (Huelsenbeck and Ronquist 2001) implemented in Geneious for 1.1 million generations, with 4 incrementally heated chains (temperature = 0.2). For the topology, an

unconstrained branch length prior was assumed, with branch length exponential prior with mean equal to 10. Markov chains were sampled every 200 generations after the first 250 000 generations. The datasets were analyzed under the model of evolution selected by MrModeltest (Nylander 2004), for both the maximum likelihood as the Bayesian sequence analysis. Generated trees were summarized with a 50% majority rule consensus tree in PAUP* (Swofford 2002) to report the posterior probabilities of each clade on the maximum likelihood tree, and was used as a measure of branch support.

4.3.6 Estimation of number of gene copies in *Oxytropis* species

4.3.6.1 Pairwise percentage divergence calculation

In each gene family analyzed here, sequences may have different type of relationships: alleles, orthologs, homeologs or paralogs. Alleles are versions of the same gene in individuals of the same species, and orthologs refer to homologous genes that arose through speciation, and that are from different species. Homeologs are found in allopolyploid genomes, and correspond to genes that were previously orthologs in ancestral species, but that are now sheltered within a same genome after allopolyploidization. Paralogs are also in one genome, polyploid or not, but arose after a duplication event. Due to the very fragmentary knowledge of both the evolutive history of *Oxytropis* polyploid taxa, and the gene set of each gene families, no attempt was made to distinguish homeologs from paralogs, and they are here globally termed “gene copies”. In order to reach our first objective, different data were taken into account: pairwise percentage divergence, fluorescence ratio from qPCR, and phylogenetic trees (previous section).

For each three gene families, uncorrected pairwise percentage divergence was calculated with MEGA (Kumar et al. 2008) or the ape R package (Paradis et al. 2004) from the full gene regions (5' UTR, exons, introns, 3' UTR) between each sequence pair with complete deletion, and values frequencies were plotted on a histogram. From diploid genomes, a set of confident alleles (sequences from different plantlets of one species positioned together in a terminal clade of the phylogenetic tree) and confident paralogs (sequences from the same plantlet positioned in different estimated gene copies groups of the phylogenetic tree) was built.

To visualize if alleles and paralogs have strikingly different divergence values in the gene families studied here, they were mapped (black and white, respectively) on the histogram. A clear gap between values from alleles and paralogs may serve as guide in setting threshold values, useful for our conservative copy number estimation. The sequence divergence in PR-10 gene family in *Betula pendula* (Schenk et al. 2006) and *Malus domestica* (Gao et al. 2005), was also considered, as well as intron size, that distinguished PR-10 paralogous genes in *Betula* (Schenk et al. 2006).

4.3.6.2 Copy number estimation from quantitative real-time PCR (qPCR)

A quantitative real-time PCR (qPCR) on total DNA was also performed to estimate the number of copies of KS-dehydrin and PR-10 in the two diploid species (*O. arctobia* and *O. splendens*), and the two polyploid species (*O. maydelliana* and *O. campestris johannensis*) relatively to a reference gene. This approach is compatible with the minute amount of total DNA yielded by each individual

Oxytropis plantlet, which is not the case with the Southern blots approach that requires micrograms of total DNA

4.3.6.2.1 Selection of reference gene

Two reference genes were tested. They are in single copy in four angiosperms species (Duarte et al. 2010) (among 395 genes); and in an algae, a moss and two angiosperm species (Armisen et al. 2008) (among 202 genes). The two genes are small nuclear ribonucleoprotein-associated protein B (AT4G18372 in *Arabidopsis thaliana*), and thylakoid lumenal 15.0 kDa protein (AT5G52970 in *A thaliana*). A BLASTP similarity search done in October 2012, in Phytozome v8.0 (Goodstein et al. 2012) and NCBI GenBank (Sayers et al. 2010) with permissive threshold (E-value of 0.1) using the *Arabidopsis* gene sequence as query to confirm that only one hit was returned for both *Medicago truncatula* and *Glycine max*.

These hits, for small nuclear ribonucleoprotein-associated protein B (abbreviated here snrnp) from *M. truncatula* (MTR_2g019100 in NCBI) and from *G. max* (LOC100792908 in NCBI); and thylakoid lumenal 15.0 kDa protein (abbreviated here tlp15) from *M. truncatula* (MTR_5g037320 in NCBI) and from *G. max* (LOC100795443 in NCBI) were used to design degenerate primers (snrnp_F_4 5' TCGGGTAMGGAAGCTGCTGT 3' with snrnp_R_3 5' TGACATGACRTGCGGCAAGA 3'; and Thyl_lum_F_5 5' TGCTCTMTCYGGTGCACTCTCAC 3' with Thyl_lum_R_3 5' CTGTCCATCGGAGAGGAAYCCA 3') with Primer3 (Rozen and Skaletsky 2000). The two genes were PCR amplified in *Oxytropis arctobia* (plantlets 43b and 46) and in *O. splendens* (plantlets 88f and 44h) using regular PCR conditions and iProof polymerase (catalog number 172-5301, Bio-Rad laboratories Canada, Mississauga, Ontario), with a final concentration of 2.7 mM $MgCl_2$, 0.2 mM dNTP and 0.4 μM

primers; as well as 0.8 units of enzyme and approximately 10 ng of DNA, in a 50 μl volume. The thermocycling program applied was: 98˚C for 30s; followed by 30 cycles of denaturation at 98˚C for 6s; annealing at 56.5˚C for 30 s; extension at 72˚C for 1 minutes; and a final extension of 7 minutes at 72˚C. The PCR reactions we visualized on agarose gels, and sent unpurified for sequencing from both ends (Sanger sequencing on a 3730xl DNA Analyzer of Applied Biosystems at the McGill University and Génome Québec Innovation Center).

4.3.6.2.2 Quantitative real-time PCR (qPCR) reactions with total DNA

Oligonucleotides (primers and probes, Supplementary Table S4.5.) for qPCR were designed with the Primer3 (Rozen and Skaletsky 2000) algorithm and were synthesized by AlphaDNA (Montréal, Canada). Probes were labeled with the fluorescent reporter dye 6-FAM on the 5' end, and the quencher dye BHQ-1 on the 3' end, as in TaqMan experiments. Oligonucleotides for the two target gene families (KS-dehydrin and PR-10) were designed to anneal all *Oxytropis* sequences of the gene family (PR-10 and KS- dehydrin), and of the two reference genes (snrnp and tlp15). Final reaction mix contained a 1X of the SsoFast Probes Supermix (BioRad cat no 172-5230), 0.1 μM of a probe, 0.9 μM of each primer, various quantities of DNA and water, in a 10 μl volume. The DNA concentration was measured on the Qubit instrument with the High Sensitivity buffer (Life Technologies). DNA quantity varied from 0.28 ng to 9 ng of DNA, in a six 1/2 serial dilutions. The qPCR experiment was carried out on a Viia7 instrument (LifeTechnologies) on a 384 plate (Axygen cat no PCR-382M2-C) sealed with optical plate cover (ProGene cat no 45-2NL). The thermocycling applied included a 2 min incubation step at 50˚C, a 5 min activation at 95˚C, and 50 cycles of 10 sec denaturation at 95˚C, and 1.5 min annealing/extension at 60˚C, where fluorescence was recorded by the instrument.

Four technical replicates were run for each sample. Initially, two different probes assays were tested for the target gene families (PR-10 and dehydrin), and two reference genes (snrnp and tlp15) were tested as well. For the final assay however, only the probe assay with the best efficiency was retained among the two for a given gene, and the two reference genes. The outliers identified by the Viia7 v 1.2 analysis software were omitted from further analyses, and a standard curve for each gene family and the reference was built.

4.3.6.2.3 qPCR data analyses for estimation of copy number

The number of copies for a target gene family is here estimated relative to an endogenous reference gene, which is presumably in single copy in most plant genomes. A calibrator sample with a known number of copies of the target genes is commonly included in other copy number variation studies (Bubner and Baldwin 2004). Here, however, the genuine number of copies for any gene, including for snrnp and tlp15, remains to be determined since no *Oxytropis* genome is yet described; and calibrator samples with known number of copies of the target genes do not exist. In cases where reactions efficiencies are different, a simple 2-delta Ct method may not be accurate, and a global fitting approach (Carr and Moore 2012) can be used to model the curve of the raw fluorescence emitted at each cycle of the qPCR reaction, and mathematically determine template abundance, which can then be compared between target and reference. This global fitting approach works without incorporating efficiency values, log transformation of data, or exclusion of baseline values (Carr and Moore 2012).

Here, standard curves and reaction efficiencies were all calculated within the Viia7 v 1.2 analysis software, for both target gene families (PR-10 and KS-dehydrin) and for reference gene, based on Ct values of the 6-fold ½ serial dilutions of total

DNA, starting with an 9 ng of DNA. The global fitting approach was used to determine the ratio of initial template abundance (i.e. copy numbers) of the target gene relative to the reference gene, from the raw fluorescence data at each of the 55 cycles of the qPCR reactions, using the pcrbatch and the ratiocalc functions with cm3 model (Carr and Moore 2012), implemented in the qpcR package (Ritz and Spiess 2008) in the R programming language. The ratiocalc function includes calculation of the standard deviations of the ratio from the available technical replicates.

4.3.6.3 Copy number estimation from phylogenetic tree topology

Although many reconciliation methods exist to infer relative timing of gene duplications and losses in relation to speciation events; the present objective is to gain an initial reasonable and conservative estimation of the number of copies in each species. Accordingly, the number of gene copies in an *Oxytropis* species was estimated in light of the phylogenetic tree topology and branch support, according to the following logic:

- The number of different non-allele sequences from one individual plantlet is the minimal number of gene copies for the species. The risk that PCR errors are still present in the analyzed sequences is minimized by the sequence assembly steps performed earlier.
- Genes from different individuals of one species that are not separated by more than one well-supported branches are considered potential alleles. These are often sisters in the phylogenetic trees.
- Pairs of sequences separated by one or more well supported branches, but with sequence divergence in the range of allelic values, are considered alleles

(for gene families where alleles and paralogs have sharp differences in divergence values).

- A set of potential alleles of a species is considered to represent one gene copy.
- For a given species, the different gene copies are separated by more than one well-supported branches.
- If the gene count clearly detected in high polyploids was not a multiple of the count in the diploid species, it was not modified, in agreement with the conservative approach.

4.4 Results

4.4.1 Diploid and polyploid *Oxytropis* species contain gene copies of PR-10, ripening-related protein and KS-dehydrin genes

Our previous transcriptome study of two arctic and two temperate *Oxytropis* species indicated that several similar copies of gene families were expressed (Archambault and Strömvik 2011). In the present study, we show the occurrence of more than one gene copy in each genome by PCR amplification from genomic DNA, followed by direct sequencing, and sequence assembly.

A total of 32 different sequences were isolated of the PR-10 genes (nine from *O. arctobia,* 6 from *O. maydelliana*, five from *O. campestris* subsp. *johannensis,* eleven from *O. splendens,* one from *O. lambertii*), 19 sequences of the ripening-related proteins (four from *O. arctobia,* six from *O. maydelliana*, four from *O. campestris* subsp. *johannensis,* four from *O. splendens*), and 30 sequences of the KS-dehydrin (eight from *O. arctobia,* seven from *O. maydelliana*, six from *O.*

campestris subsp. *johannensis,* eight from *O. splendens;* one from *O. lambertii*). We also isolate eight sequences of the low-copy genes *lhcaIII* (three from *O. arctobia,* one from *O. maydelliana*, two from *O. campestris* subsp. *johannensis,* two from *O. splendens*) and eight of the *lhcbI* (two from *O. arctobia,* two from *O. maydelliana*, one from *O. campestris* subsp. *johannensis,* two from *O. splendens*; one from *O. lambertii*). All sequences are deposited in GenBank under accession numbers HQ731797 to HQ731908.

4.4.2 Gene structure is conserved for the genes surveyed among species and among gene copies

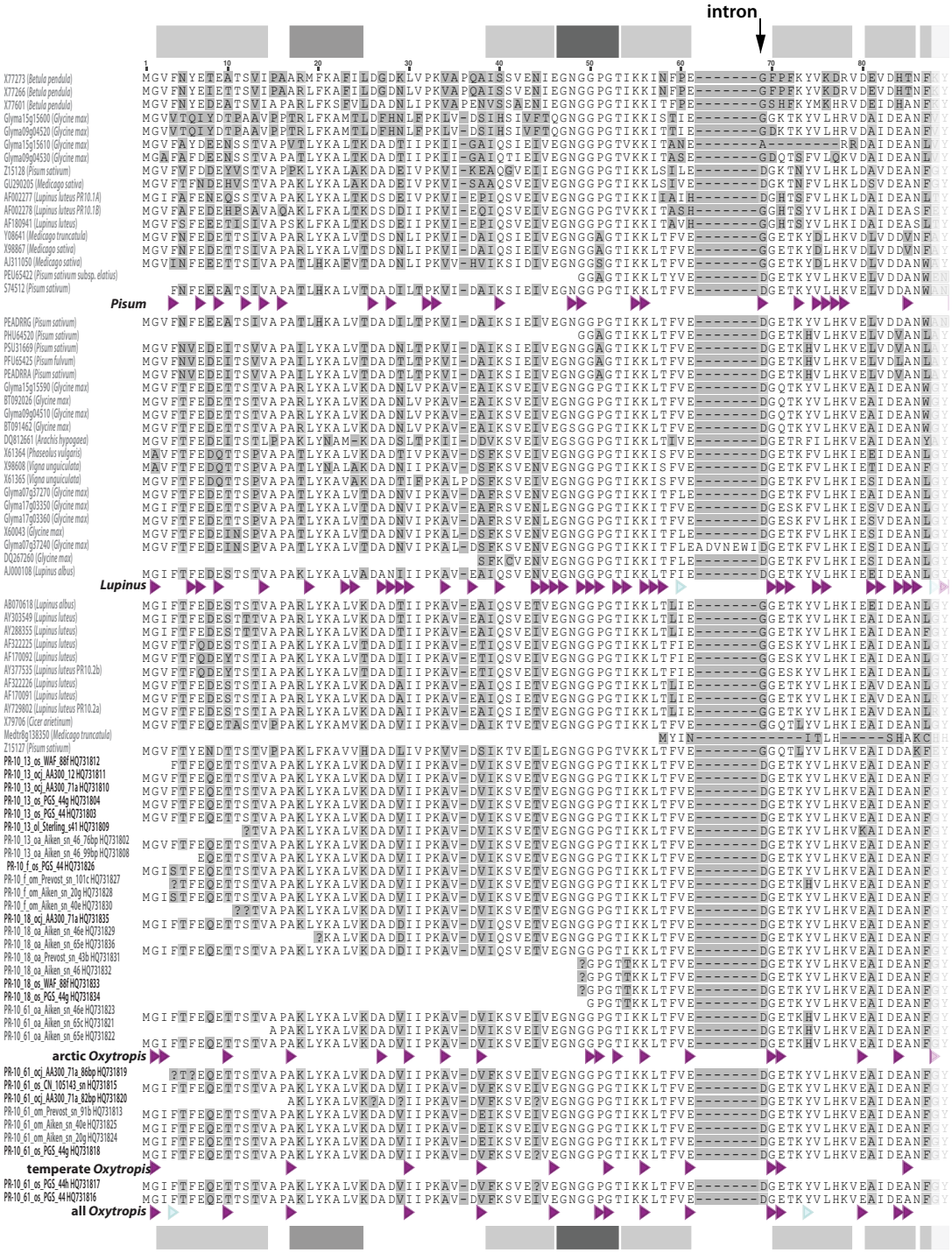
To locate the coding sequence region, the start codon, intron and exon boundaries and stop codons in *Oxytropis* genes were inferred from described Fabaceae genes. The results indicate that the single intron is similarly located in both PR-10 and ripening-related *Oxytropis* gene families: it starts at nucleotide position 180 and at position 183 or 186, respectively (Fig. 4.1a, b). Intron location is also conserved when compared to similar PR-10 from *Lupinus luteus* (Ypr10.2b AY377535 gi:34978686 and Ypr-10.2a AY729802 gi:52352967) and ripening-related proteins from *Medicago truncatula* (mRNA BT053279 gi:217075157 and portion of a genomic clone AC146758). The *Arabidopsis* (a Brassicaceae) *MLP28* gene (AT1G70830), more distantly related to ripening-related proteins, has, however, two introns. The PR-10 intron is generally shorter than the ripening-related intron (Supplementary Table S4.3).

The *Oxytropis* KS-dehydrin genes (Fig. 4.1c) all contain an intron that varies in length (Supplementary Table S4.3) in the 3' UTR region (not shown), as in similar dehydrins (e.g. *Medicago truncatula* MtCAS15, EU139869 gi:161897790).

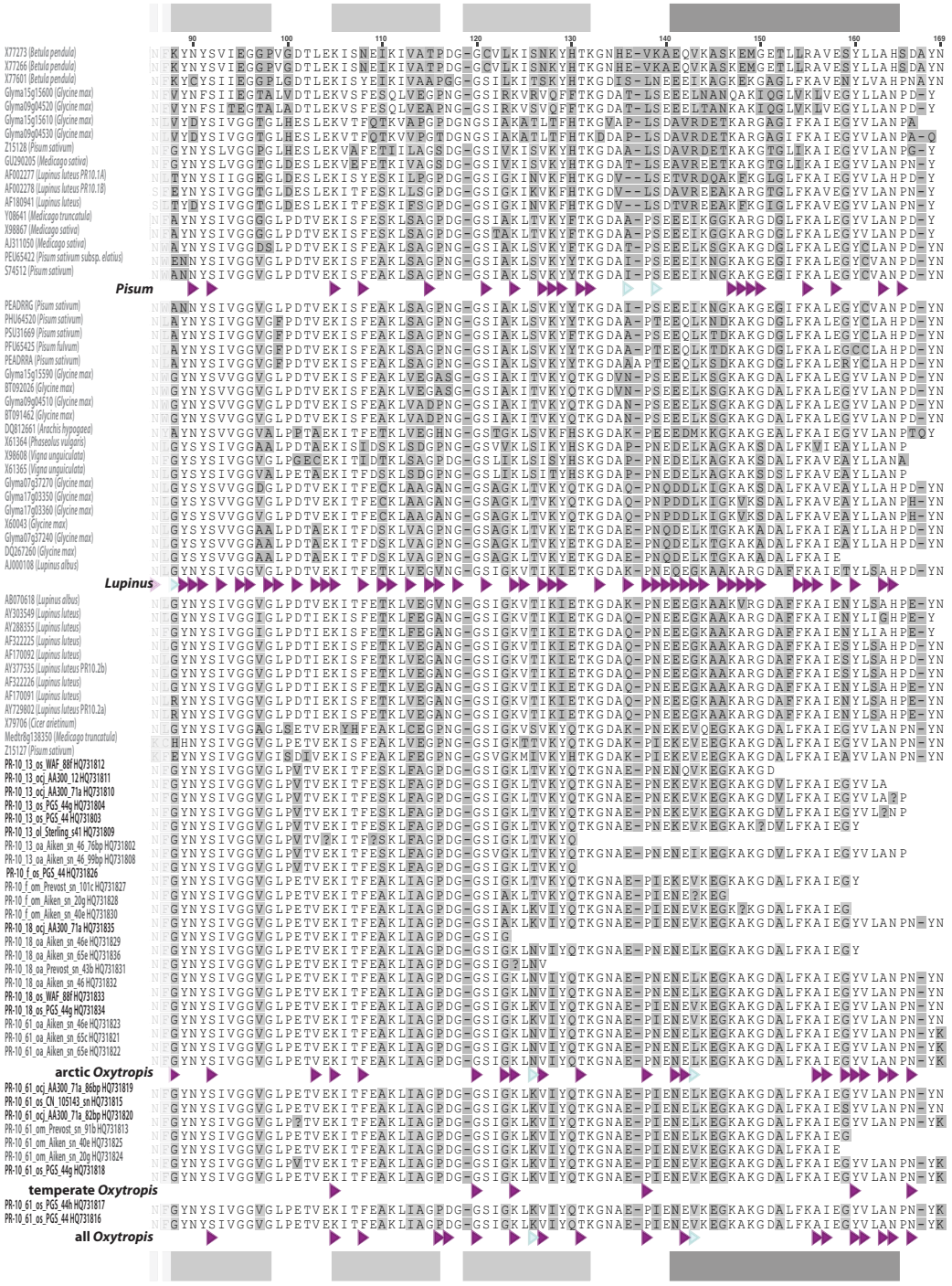
The light-harvesting complex A III gene (*lhcaIII*) coding sequence is interrupted by two introns (Supplementary Table S4.3), separated by a 68 bp exon, as is *lhcaIII* of *Arabidopsis* (AT1G61520). The first intron is short while the second is longer, except for the *O. splendens* plantlet 23d where the second intron is shorter than in other *Oxytropis* species (Supplementary Table S4.3). The *Oxytropis* light-harvesting complex B I gene (*lhcbI*) amplified from genomic DNA are uninterrupted by introns, as is the *Glycine max lhcbI* gene (GMU01964 gi:506628). The alignments therefore show that, when present, intron location is conserved among genes from *Oxytropis* and other Fabaceae and Brassicaceae (*Arabidopsis*) species.

Figure 4.1 Amino acid alignment of *Oxytropis* sp. genes, along with similar Fabaceae sequences

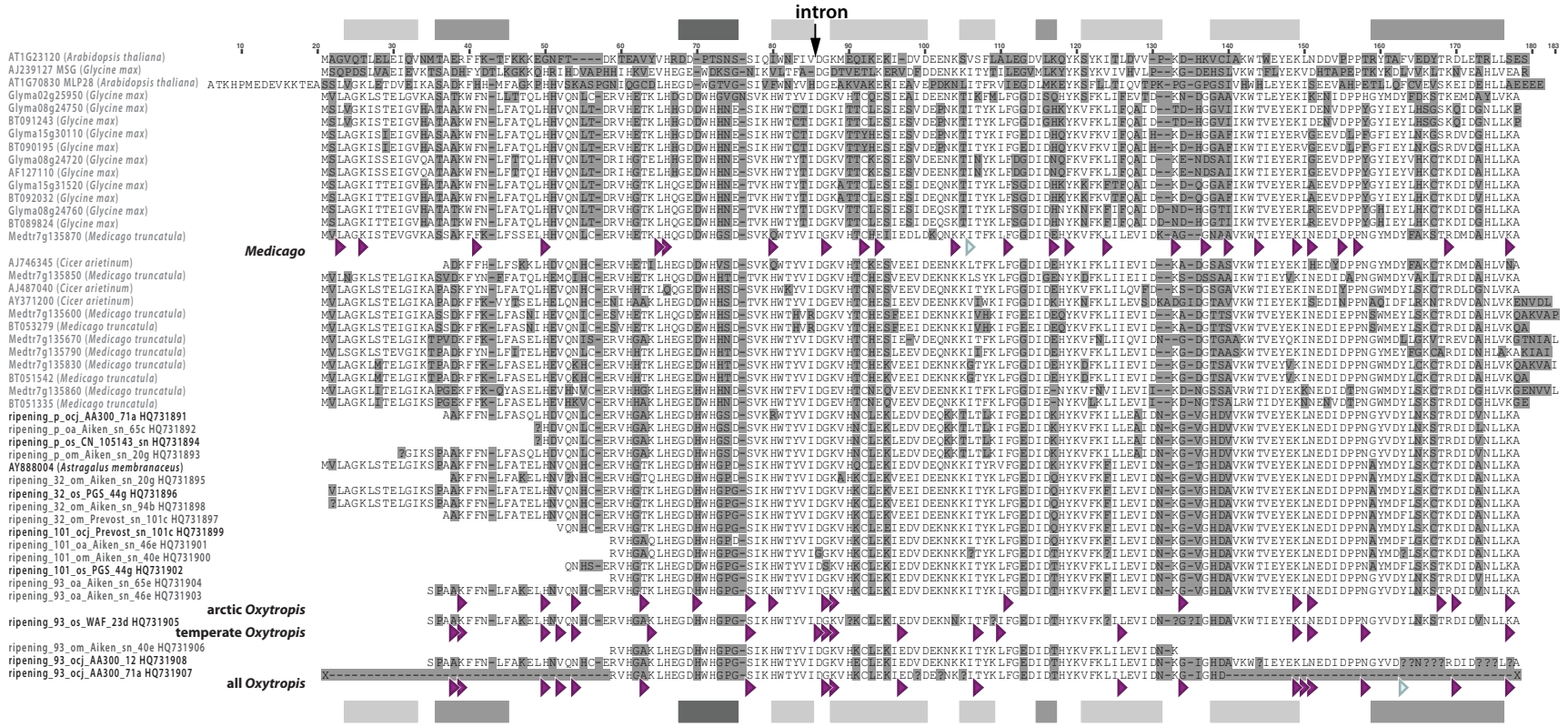
Amino acid sequences were conceptually translated from the coding region of sequenced fragments from genomic DNA. Aligned residues with lower than 60% similarity are highlighted in grey. Codons under negative selection are identified by a dark triangle and codons under positive selection by a light grey triangle. For the analysis comprising all *Oxytropis* sequences, triangles are shown below the bottom *Oxytropis* sequence in the alignment; for the separate arctic or temperate sequences analysis, triangles are below an *Oxytropis* sequence of the appropriate species, and for the analyses of other Fabaceae genera, triangles are below a representative species sequence. Secondary structures for similar proteins with an experimentally determined structure, or other features such as repeats, are indicated in the alignment by underlying grey boxes. **a)** Alignment of the PR-10 proteins and **b)** alignment of the ripening-related proteins. Predicted ß sheets are in light grey, α helices in medium grey, and the glycine-rich region in darker grey. **c)** Alignment of the KS-dehydrins. Repeats similar to the Y-segment are in light grey, B-repeats in medium grey, the two HEHG repeats in darker grey. The K-segment and S-segment are highlighted by underlying near black boxes. Name for *Oxytropis* sequence refer to: gene family, gene copy, species, collector, collection number, plantlet number, and when two sequences need to be further differentiated, length of intron and additional copies within a clade is included in the name.



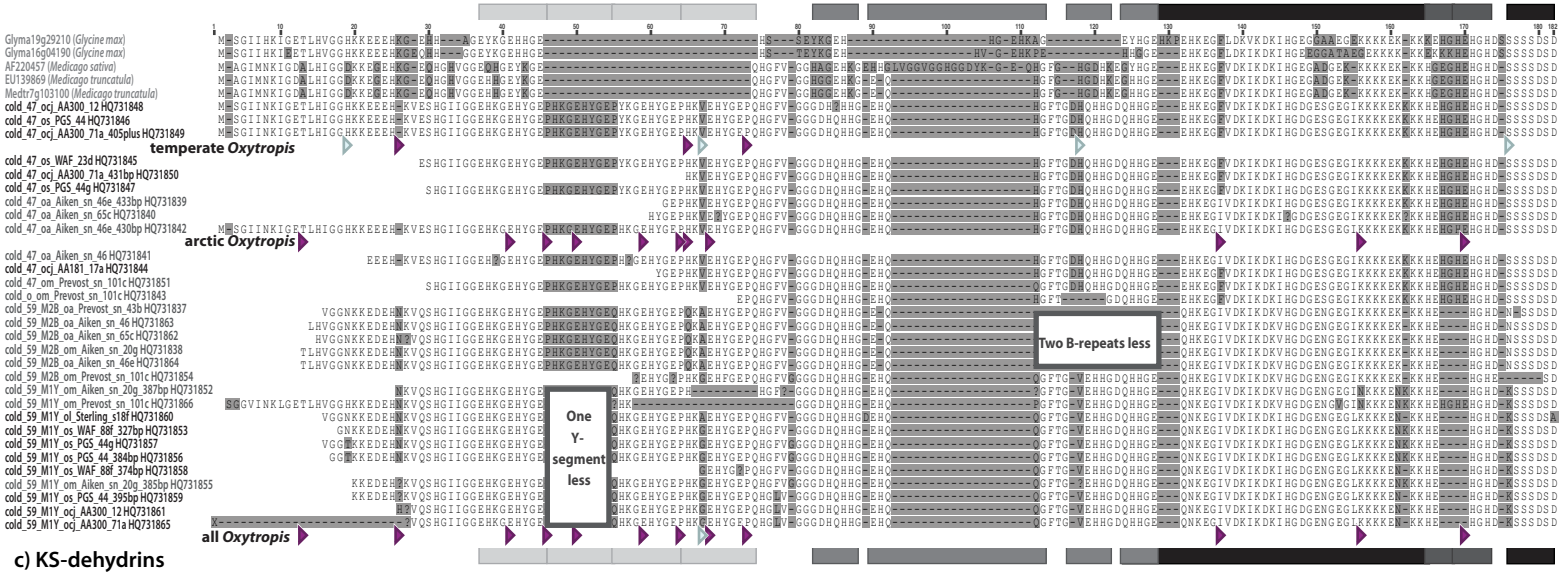
a) Pathogenesis related class 10 (PR-10)



a) Pathogenesis related class 10 (PR-10)



b) Ripening-related proteins



c) KS-dehydrins

4.4.3 Codons evolving under positive and negative selection are detected in different proportions in genes from arctic and temperate *Oxytropis* species

In order to gain a better understanding of the selection that acted on the coding sequences of the members of PR-10, ripening-related and KS-dehydrin families, as well as two low-copy *lhcaIII* and *lhcbI* genes, codons under positive and negative selection were detected (Table 4.1) and located (Fig. 4.1) in *Oxytropis* arctic and temperate species. Negative selection suggests that codons were under conformational and functional constraints; and positive selection that amino acid changes were once advantageous at a site. This will bring us better knowledge of proteins from plants adapted to the Arctic.

Preventive measures were taken to minimize the negative effect recombinant sequences may have on selection detection. The processes generating recombination (genetic recombination or artifact chimeras) are, however, not distinguished from one another here. One analysis revealed that recombination was suspected in two gene families, where the breakpoint is located within the second exon of the PR-10 and of the ripening-related proteins genes; and in *lhcaIII* (Table 4.1). Another test and a careful visual inspection of the sequence alignment revealed that only one sequence (the PR-10 HQ731814, from *O. splendens* plantlet 88f) might be a PCR-induced chimera. This sequence was excluded from all sequence analyses.

Generally, codons under negative selection are well distributed along the gene sequences (Fig. 4.1), except for the KS-dehydrins, where they cluster in the Y-segment (see below, Fig. 4.1c) for the analyses of arctic and of all *Oxytropis* sequences. Given that the genes surveyed here were differentially expressed in arctic and in temperate *Oxytropis* plantlets (Archambault and Strömvik 2011), they may

be subjected to different selective pressure at the nucleotide sequence level in the arctic or in the temperate species, and this is addressed by our second question (Do PR-10 and KS-dehydrin genes, found overexpressed in arctic species, have different selective pressure in arctic species, compared to temperate species?). Detection of codon under selection was therefore performed for all *Oxytropis* sequences, but also for sets partitioned by geographic origin. When arctic and temperate *Oxytropis* sequences are analyzed separately, the proportion of codons under negative selection is similar for ripening-related proteins, but is higher in the arctic set for PR-10 and KS-dehydrin (Table 4.1). The *Lupinus* PR-10 set of sequences stands out for its elevated proportion of negatively selected codons, whereas this proportion is markedly low in the set of KS-dehydrins from temperate species. Codons under positive selection are overall rare or absent in the surveyed genes. They are more common only in the full set *Oxytropis* PR-10 (Table 4.1), that includes arctic and the temperate sequences, where they are dispersed along the sequence (Fig. 4.1b).

Table 4.1 Proportion of codons under positive and negative selection in genes or gene families from four *Oxytropis* species genomes.

Gene (or gene family)	Number of codons in alignment	Number of sequences	Recombination suspected	Number of codons under positive selection (% of alignment)	Number of codons under negative selection (% of alignment)
PR-10					
Complete *Oxytropis* set	158	32	Yes, at 104 aa	4 (2.53)	33 (20.9)
Arctic *Oxytropis* set	158	16	None	2 (1.27)	38 (24.1)
Temperate *Oxytropis* set	158	16	Yes, at 101 aa	0	17 (10.8)
Lupinus set	158	13	Yes at 18 aa	2 (1.27)	84 (53.2)
Pisum set	159	9	Yes, at 73 aa	2 (1.26)	42 (26.4)
Ripening-related					
Complete *Oxytropis* set	152	18	Yes, at 90 aa	1 (0.65)	19 (12.5)
Arctic *Oxytropis* set	152	10	Yes, at 128 aa	0	17 (11.2)
Temperate *Oxytropis* set	152	8	Yes, at 144 aa	0	18 (11.8)
Medicago set	157	10	Yes at 100 aa	1 (0.64)	24 (15.3)
KS-dehydrin					
Complete *Oxytropis* set	158	30	None	1 (0.63)	10 (6.33) [a]
Arctic *Oxytropis* set	157	15	None	0	10 (6.37) [a]
Temperate *Oxytropis* set	158	15	None	3 (1.90) [a]	2 (1.27) [a]
lhcaIII; *Oxytropis* set	203	8	Yes, at 118 aa	0	14 (6.89)
lhcbI; *Oxytropis* set	240	8	None	0	24 (10.0)

[a] Excluding sites upstream codon position 27 in the alignment

aa: Amino acid position in sequence alignment

4.4.4 Likely secondary structures are estimated for *Oxytropis* PR-10, ripening-related and light-harvesting proteins based on similarity to proteins with determined structures

In order to understand the detected codons under selection in the context of protein secondary structure, and to answer our third question on similarities and differences between the related families PR-10 and ripening-related proteins, a likely protein structure was inferred for *Oxytropis* genes, using similar genes where protein structures are experimentally determined. Such proteins, very similar to LHCAIII, LHCBI and PR-10 genes, are identified from angiosperms species, as show by the low *E* value (Table 4.2). A ripening-related protein from the Brassicaceae with a determined structure is similar to the *Oxytropis* gene, albeit to a lesser extent (Table 4.2). Protein ß sheet and α helix secondary structures from non-*Oxytropis* genes were then mapped in the amino acid sequence alignment (Fig. 4.1). The few positively selected codons in PR-10 are mainly located in ß sheets. Since characterized dehydrins show properties of intrinsically disordered proteins (Mouillon et al. 2008), no experimentally determined protein structure was linked with these proteins.

Table 4.2 Proteins with experimentally determined structure that are similar to *Oxytropis* sequences for the PR-10, the ripening-related proteins, and the light-harvesting genes.

Oxytropis sequence query	Similar protein from RCSB PDB archives[1]	Gene name	Plant species	*E* value	Reference
PR-10 "paralog 61"	2QIM	LlPR-10.2B in complex with cytokinin	*Lupinus luteus*	2.86 E-63	(Fernandes et al. 2008)
PR-10 "paralog 18"	2K7H	Gly m 4	*Glycine max*	9.48 E-53	(Berkner et al. 2009)
PR-10 "paralog 13"	2K7H	Gly m 4	*Glycine max*	4.73 E-63	(Berkner et al. 2009)
Ripening-related proteins "paralog 93"	2I9Y	At1g70830	*Arabidopsis thaliana*	2.86 E-15	(Lytle et al. 2009)
Ripening-related proteins "paralog A"	2I9Y	At1g70830	*Arabidopsis thaliana*	3.71 E-13	(Lytle et al. 2009)
Ripening-related proteins "paralog 32"	2I9Y	At1g70830	*Arabidopsis thaliana*	1.93 E-20	(Lytle et al. 2009)
Ripening-related proteins "paralog 101"	2I9Y	At1g70830	*Arabidopsis thaliana*	4.82 E-14	(Lytle et al. 2009)
LHCAIII	2WSC	Plant Photosystem I	*Pisum sativum*	4.41 E-89	(Amunts et al. 2010)
LHCBI	1VCR	Plant Photosystem II	*Pisum sativum*	8.15 E-130	(Hino et al. 2004)

[1](Berman et al. 2000)

4.4.5 The *Oxytropis* dehydrin sequences have a K-like - Y4 – K– S structure, show novel repeats and their Y-segment is conserved

Although they are intrinsically unstructured, dehydrins are characterized by repeats that resemble functional sites (Close 1996). The *Oxytropis* sequences were therefore inspected for presence of known (K-segment, Y-segment, S-segment) or novel repeats. The amino acid sequence alignment indicates that *Oxytropis* dehydrins (Fig. 4.1c) have four Y-segment repeats (HKGEHYGEP) in the same position as similar genes (*Medicago* BudCAR5 and MtCAS15). Glutamate (E) in all four repeats and proline (P) in two of the four repeats are under negative selection (Fig. 4.1c) in the arctic sequences, but in the temperate sequences it is another proline (P) that is under negative selection. Within the last repeat, one codon (G/A/V) is positively selected in the temperate sequences. Eleven *Oxytropis* sequences (arctic *O. maydelliana* HQ731852, HQ731866, HQ731855; temperate HQ731860, HQ731853, HQ731857, HQ731856, HQ731858, HQ731859 HQ731861, HQ731865) lack a repeat (indicated in Fig. 4.1c and Fig. 4.3c by the "One Y-segment less" box) and two *O. maydelliana* sequences (HQ731852, HQ731866) have additional 9 and 20 amino acids deletion in this region, reducing the number of repeats to two and one in these cases. Another type of repeat, here called B-repeats (DHQHHG), is downstream of the Y-segment, and is repeated three times except in some *O. maydelliana* and *O. arctobia* sequences (HQ731837, HQ731863, HQ731862, HQ731838, HQ731864) where only one unit is present (indicated in Fig. 4.1c and Fig. 4.3c by the "Two B-repeats less" box). One codon (V/G/H) of a B-repeat is under positive selection in the temperate species, but not in the arctic ones. The *Oxytropis* sequences show known K-segment variants (HKEGFVDKIKDKIHG and NKEGIVDKIKDKVHG), a K-rich segment (KKKKEKKKK) and, in the C-

terminal region, a dehydrin S-segment (SSSSDSD) (Battaglia et al. 2008). Although the amino acid sequence is conserved in the sequence alignment in these three regions, they are not under stringent selection because no or very few codons are under negative selection (Fig. 4.1c). The first codon (S/N/K) of the short S-segment is under positive selection. One *O. maydelliana* sequence (HQ731854) lacks the S-segment. Most of the *Oxytropis* dehydrins have a K-like segment- Y_4 segment – K segment – S segment structure, and were therefore termed KS-dehydrin. The overall amino acid composition is biased with 19.3% glycine (G), 17.6% histidine (H), 13.5% glutamate (E) and 13.1% lysine (K).

4.4.6 There are at least three PR-10 and ripening-related protein genes; and two KS-dehydrin genes in *Oxytropis* diploids

Particular attention was paid to detect gene copies, since some of the *Oxytropis* species *(O. maydelliana* and *O. campestris* subsp. *johannensis)* are polyploid, and some of the surveyed genes are arranged into gene families in other plants. To minimize PCR-induced errors in gene sequence, the strategy adopted here assembled commonly five and sometimes up to ten sequences, that were carefully inspected. As mentioned, sequence pairs in this dataset may have different type of relationships: alleles, orthologs, homeologs or paralogs. Due to the very fragmentary knowledge of the evolutive history of *Oxytropis* polyploid taxa, and the likely incompleteness of the datasets; attempt to distinguish homeologs from paralogs in polyploid genomes may be hazardous, and these are consequently globally termed “gene copies”.

The unusually high number of double peaks in the chromatogram for the amplified low-copy photosynthesis genes (*lhcaIII* and *lhcbI*) from the *O.*

maydelliana polyploid is one observation that suggests occurrence of very similar gene copies (homeologs or paralogs) of same length in this species. To better answer our first question on the occurrence of more than one copy in *Oxytropis* genomes, three types of results were considered, sequence divergence, fluorescence ratio of the target gene families copies over a reference gene (qPCR), and sequences evolutionary relationships.

4.4.6.1 Distribution of sequence divergence values

The distribution of sequence divergence values (Supplementary Table S4.4, Fig. 4.2) was inspected to detect presence of more than one distinct mode, in which case the group of sequence pairs with low sequence divergence values may be alleles (black dots in Fig. 4.2), and the group with higher values may be gene copies (white dots in Fig. 4.2).

For the ripening-related proteins, the sequence divergence clearly appears distributed into two groups (Fig. 4.2b), one with very low values (0%), and one with high values (19% to 26%). The distribution of divergence values for KS-dehydrin (Fig. 4.2c) also suggests more than one mode, but more similar in values, one with low divergence values (0% to 1.7%), and the other one or two modes with slightly higher values (5.2% to 8.7%). The distribution of pairwise divergence in PR-10 sequences (Fig. 4.2a) is more difficult to interpret as no clear gap in sequence divergence values exists. PR-10 sequence pairs that may correspond to alleles diverge by up to 1.3%, similar to values in *Betula pendula* (Schenk et al. 2006) and *Malus domestica* (Gao et al. 2005) PR-10; while pairs of gene copies may have values between 5.4% to 8.5%.

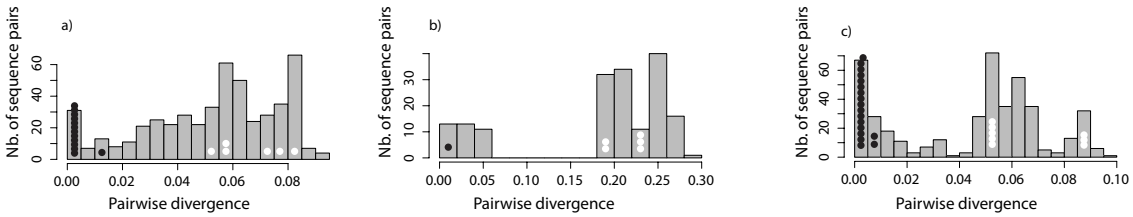


Figure 4.2 Pairs of confident alleles and confident paralogs displayed on a histogram showing the distribution of pairwise sequence divergence values for three gene families surveyed, in four *Oxytropis* species.

a) The PR-10 proteins on 224 positions, where no clear gap in divergence values exists between the confident alleles and the confident gene copies; **b)** The ripening-related proteins on 289 positions, where the X axis shows higher upper values; and **c)** the KS-dehydrins gene families on 230 positions. Uncorrected sequence percentage divergence was calculated for each pairs of sequences, with complete gap deletion, considering the whole gene sequence (including CDS, intron, 5' and 3' UTR). From diploid genomes, the pairwise divergence for confident allele pairs is reported by black dots, and for confident paralogs pairs by white dots.

4.4.6.2 Estimation by qPCR supports occurrence in multiple copies

Distribution of sequence divergence values in the PR-10, ripening-related proteins and KS-dehydrins gene families thus suggests occurrence of gene copies, but does not inform on a possible number of copies in each *Oxytropis* genome. To this end, a qPCR experiment was performed using primers and probes annealing to all known *Oxytropis* gene copies, for two target gene families (PR-10 and dehydrin) and for a reference gene (small nuclear ribonucleoprotein-associated protein B, snrnp). Reactions' efficiencies were different enough (Supplementary Table S4.6)

between genes and between genomes, to warn against using a simple 2-delta Ct comparison of the target and reference genes as indication of the relative copy number in each genome. Instead, a mechanistic "global fitting" approach (Carr and Moore 2012) was favoured.

The ratio for the KS-dehydrin copy number compared to the reference snrnp is approximately 1 (Table 4.3) in all but one of the *O. maydelliana* plantlet. If the reference gene were assumed to be in one copy in diploids, this would imply a single copy of the KS-dehydrin in diploids as well. However, since sequence data obtained here point to occurrence of at least two copies of dehydrin (Fig 4.3), the snrnp reference gene is also likely in two copies in *Oxytropis* diploids genomes. Fluorescence ratios from the qPCR experiment suggest that PR-10 genes occur in variable number of copies among the different *Oxytropis* species, and also between individuals of a species, in contrast to the KS-dehydrin where the ratio is stable. In the diploid species, the fluorescence was between 3.6 to 6.7 fold higher for PR-10 than for the reference snrnp (Table 4.3), which may occur in two copies as mentioned. In the hexaploid, that ratio was between 7.4 to 9.4; while it was as high as 9.0 to 12.7 in the dodecaploid (Table 4.3). It should furthermore be warned that, as copy numbers increase to more than 3 fold higher the reference gene, confidence in the ratio value progressively decreases (see for instance, Bubner and Baldwin 2004). Although the high relative copy number for PR-10 revealed by the qPCR experiment should not be strictly interpreted; it suggests a likely occurrence of more than four PR-10 copies in the diploids *O. arctobia* and *O. splendens*, more than seven copies in *O. campestris* subsp. *johannensis*, and more than nine copies in *O. maydelliana*.

Table 4.3 Estimation of copy number of PR-10 and KS-dehydrin gene families in *Oxytropis* species relatively to a reference gene[a], based on qPCR fluorescence data analyzed with a mechanistic "global fitting" approach[b].

Gene family	Ratio of initial number of gene copies over a reference gene[a] for *Oxytropis* species (standard deviations)[c]							
	Oa (plantlet 65b)	Oa (plantlet 57a)	Om (plantlet 101b)	Om (plantlet 47b)	Ocj (plantlet 87a)	Ocj (plantlet 98a)	Os (plantlet 99a)	Os (plantlet 54a)
PR-10	4.8 to 4.9 (0.6)	3.6 to 3.9 (0.7)	9.0 to 10.8 (0.9)	10.6 to 12.7 (3.9)	7.4 to 7.8 (1.7)	7.7 to 9.4 (1.8)	4.2 to 5.3 (0.7)	5.5 to 6.7 (0.6)
KS-dehydrin	1.2 to 1.5 (0.2)	1.0 to 1.2 (0.3)	1.0 to 1.1 (0.13)	1.7 to 2.4 (0.4)	0.98 to 1.1 (0.2)	0.82 to 0.95 (0.17)	0.94 to 1.0 (0.13)	0.95 to 0.98 (0.16)

[a] The reference gene was small nuclear ribonucleoprotein-associated protein B (abbreviated here snrnp).

[b] To calculate the ratio, the initial abundance (copy number) present in the qPCR reactions is determined by the mechanistic "global fitting" approach (Carr and Moore 2012) for target gene families and for the reference gene. The ratio between the two values is then calculated for the available comparisons among the reactions using 9 ng, 4 ng, 2.25 ng or 1.125 ng of template DNA. Only the higher and lower ratios are reported here, for brevity.

[c] Only the highest of the standard deviation values is reported, for brevity.

Oa: *Oxytropis arctobia*; Om: *O. maydelliana*; Ocj: *O. campestris* subsp. *johannensis*; Os: *O. splendens.*

4.4.6.3 Phylogenetic trees topology also suggests gene copies

Phylogenetic analyses (Fig. 4.3) were performed to clarify the evolutionary relationships of *Oxytropis* gene copies between themselves and in relation to other Fabaceae species.

In the phylogenetic tree from PR-10 genes reconstructed with a GTR+I+G model of sequence evolution (Fig. 4.3a), all Fabaceae sequences form a highly supported clade and all *Oxytropis* sequences form a clade also highly supported. Within that clade, four groups are observed. Genomes of the four *Oxytropis* species possess the estimated gene copy "61" and "13". Gene copy "f" is so far found only in *O. maydelliana* while three species have gene copy "18" but not *O. maydelliana*. Only two sequence pairs isolated from one plantlet are part of a same estimated gene copy. For these, the sequence divergence and the intron size were taken into account for guidance in their classification as alleles or as gene copies. The two *O. campestris* subsp. *johannensis* sequences of plantlet 71a (HQ731820 and HQ731819) are very similar (0% divergence, Supplementary Table S4.4) and are likely alleles. By contrast, the two *O. arctobia* sequences of plantlet 46 (HQ731808 and HQ731802) might consist of two gene copies based on their different intron size (Supplementary Table S4.3); but their sequence divergence is, although slightly high (2.23%, Supplementary Table S4.4), lower than PR-10 confident paralogs. This pair is consequently not counted as different genes (Table 4.4), in agreement with our conservative approach. Following the conservative steps that take into account topology, three copies of PR-10 are counted in all genomes, and four in *O. splendens*. These estimates are below the qPCR results.

In the phylogenetic tree that includes the *Glycine max MSG* (AJ239127) (Strömvik et al. 1999); the *Arabidopsis thaliana MLP28* (AT1G70830) (Lytle et al. 2009) and the core Fabaceae ripening-related genes reconstructed with a GTR+G model of sequence evolution, the latter sequences form a clade separated from *MSG* and *MLP28* by a long and highly supported branch (Fig. 4.3b). Within the Fabaceae ripening-related sequences clade, a clade is formed by the *Oxytropis* and *Astragalus* sequences, where the single *Astragalus* sequence is nested within the clade that otherwise includes only *Oxytropis* sequences. The *Oxytropis* sequences are arranged in four groups. Gene copies "A", "101" and "93" are present in the genome of each four *Oxytropis* species surveyed, but there is evidence for gene copy "32" only in the *O. maydelliana* and *O. splendens* genomes. No ripening-related pair of sequences, isolated from one plantlet are suspected to be alleles based on a low sequence divergence (Supplementary Table S4.4). Following the logical steps, three copies are counted in the diploid species and four in the polyploid species (Table 4.4).

In the phylogenetic tree for the KS-dehydrin genes reconstructed with a F81+G model of sequence evolution (Fig. 4.3c), *Oxytropis* sequences form a clade separated from the other Fabaceae sequences by a long and highly supported branch. Within the *Oxytropis* clade, three informal groups are observed. One group is the paralog "47", present in the genome of the four *Oxytropis* species surveyed. Another group is paralog "59", itself divided in two subgroups, each characterized by different deletions of amino acids repeat motifs. Sequences of the subgroup "M2B" miss at least two B-repeats (Fig. 4.1c); and sequences of the "M1Y" miss at least one Y-segment, compared to the others *Oxytropis* sequences (Fig. 4.1c). Our results provide evidence for "M1Y" gene in *O. maydelliana,* in *O. campestris* subsp. *johannensis* and in *O. splendens*; but a "M2B" gene is so far only found in the arctic *O. arctobia* and *O. maydelliana* genomes. Within the larger group paralog "59", an

orphan *O. maydelliana* sequence misses the C-terminal S-segment. A few pairs of sequences isolated from one plantlet are located in a same clade, and may represent gene copies or alleles. For these, sequence divergence and intron size were be taken into account for guidance in their classification. Pair of sequences from *O. campestris* subsp. *johannensis* plantlet 71a (HQ731849 and HQ731850); from *O. splendens* plantlets 88f (HQ731853 and HQ731858) and plantlet 44 (HQ731856 and HQ731859) are very similar (Supplementary Table S4.4) and may be alleles. For another sequence pair (HQ731839 and HQ731842 from *O. arctobia* plantlet 46e), the sequence divergence is the high range for alleles (1.7%, Supplementary Table S4.4), but it is still lower than confident paralogs, and the two sequences are not counted as gene copies, according to our conservative logic. The same treatment is applied to two *O. maydelliana* sequences (HQ731851 and HQ731843) that differ by a 18 bp indel, but are very similar (0.9%). By contrast, other *O. maydelliana* sequences (HQ731854 and HQ731866, from plantlet 101c) are quite similar (1,3%, Supplementary Table S4.4), but are counted as gene copies due their many indel differences. This conclusion also applies to the *O. maydelliana* sequences that are slightly less divergent than confident paralogs (HQ731852 and HQ731855, 3.5%, Supplementary Table S4.4), but show indels difference. Following the logical steps, two copies are initially identified in all species, except *O. maydelliana* where up to five copies may coexist (Table 4.4). The qPCR experiment supports the finding of two copies in most *Oxytropis* genome, assuming the snrnp reference gene also occurs in two copies.

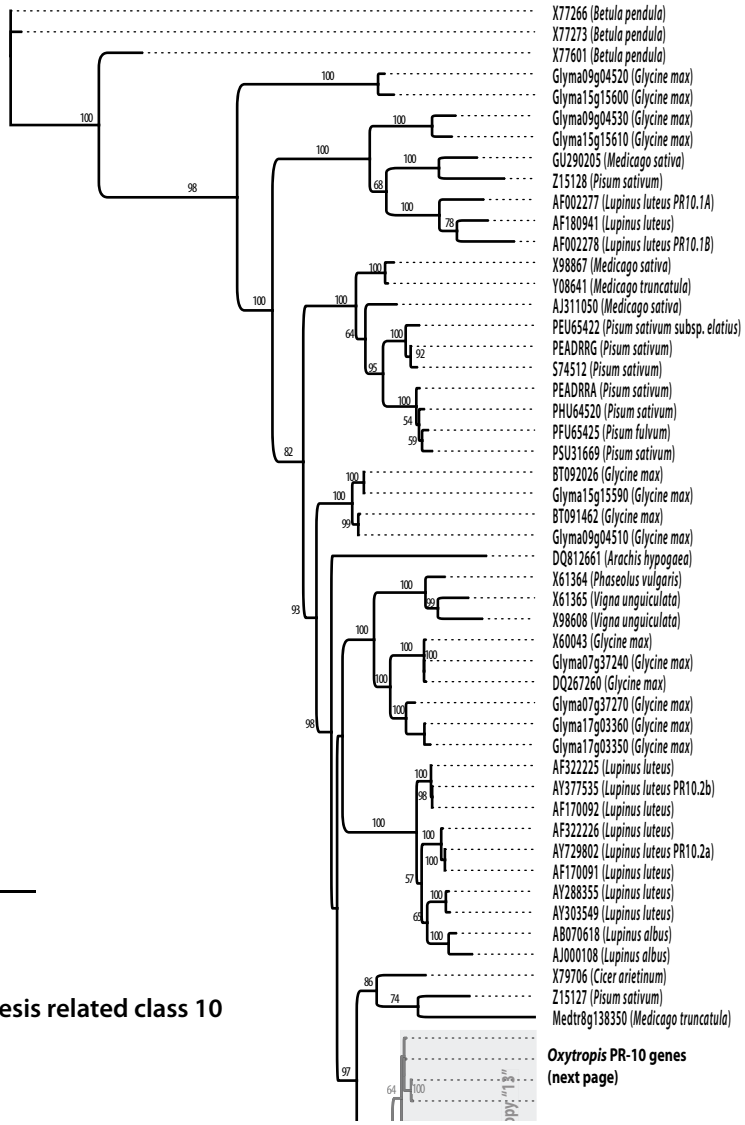
The very conservative estimation of number of copies from phylogenetic analyses suggests that the two diploid species (*O. arctobia* and *O. splendens*) and the hexaploid *O. campestris* subsp*. johannensis* have a similar number of copies for the three gene families surveyed (Table 4.4). The more sensitive qPCR experiment suggests, however, a higher and variable but still imprecise PR-10 copy number. In

some plantlets of the dodecaploid *O. maydelliana,* both phylogenetic analysis (Table 4.4), and qPCR ratios (Table 4.3) provide evidence for a slightly higher number of copies for dehydrin and for PR-10, and more comprehensive genomic searches would likely detect additional ones.

Importantly, for the three gene families surveyed, the tree topology (Fig. 4.3) shows that *Oxytropis* genes cluster in monophyletic or near-monophyletic groups, with other Fabaceae' sequences (e.g. from *Medicago* and *Glycine*) as sister group. This topology cannot fully resolve the timing of the duplications, but suggests that they predate arctic species origin; otherwise, all gene copies from a given arctic species would group into a clade. This topology also suggests that duplications happened after evolution of the lineages leading to *Oxytropis* and to *Medicago*, two legume genera only moderately related, otherwise many clades would comprise gene from both genera.

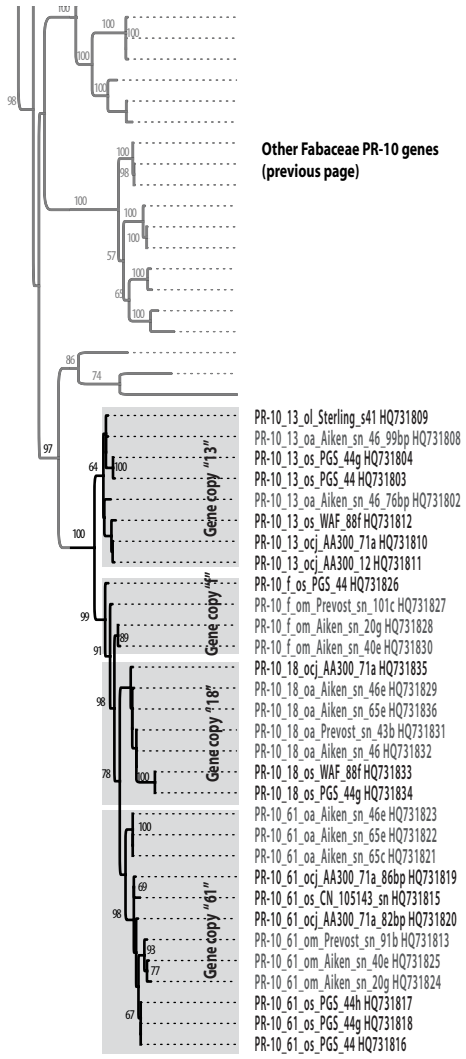
Figure 4.3 Phylogenetic trees showing relationships among several *Oxytropis* sp. genes, along with similar Fabaceae sequences

a) The PR-10 proteins b) The ripening-related proteins and c) the KS-dehydrins gene families. Trees were calculated from nucleotide sequences of the coding region (CDS) with the PHYML algorithm. Numbers above branches report Bayesian support, in the form of percentage of trees from Bayesian analysis with this branch. Scale bars denote number of substitutions per nucleotide according to the PHYML analysis. *Oxytropis* names in dark grey are from a temperate species, and names in a light grey are from an arctic species. Outgroups (in lighter grey) are all from the Fabaceae family, except At1G23120 and At1G70830 which are from *Arabidopsis thaliana*. GenBank accession number is indicated after the sequence name. Estimated paralogs are highlighted by light grey boxes. Other informative sequence features are indicated by darker grey boxes.

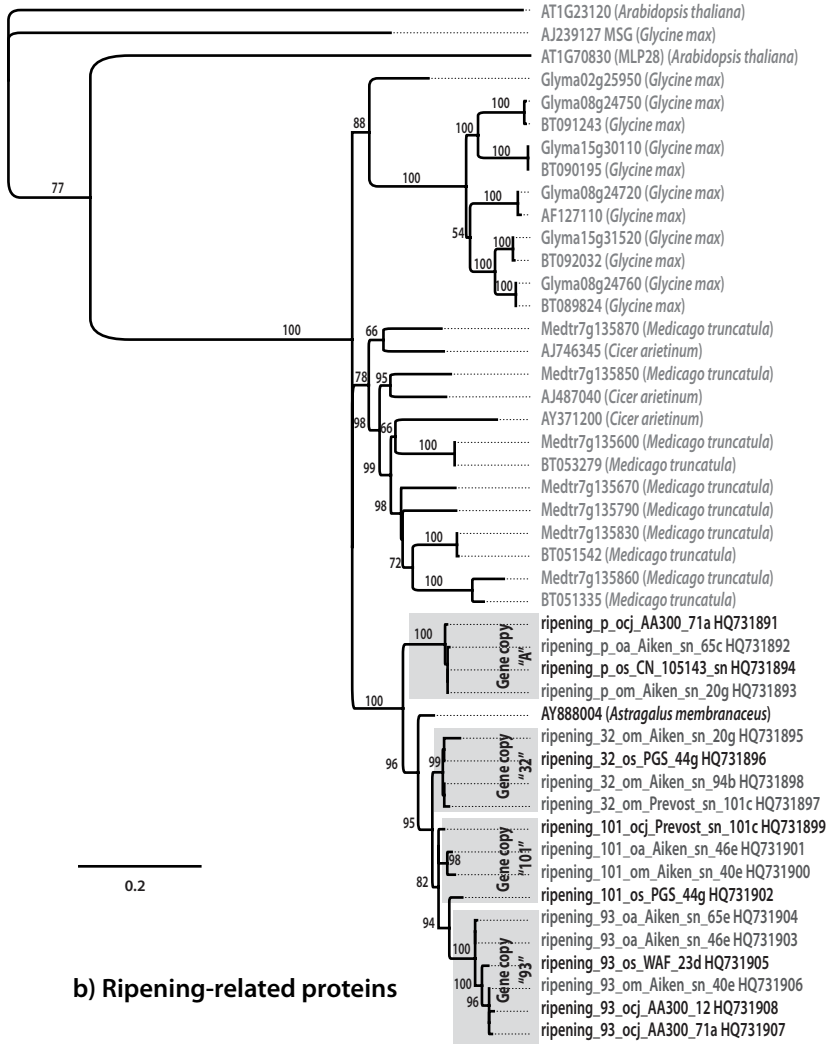


a) Pathogenesis related class 10 (PR-10)

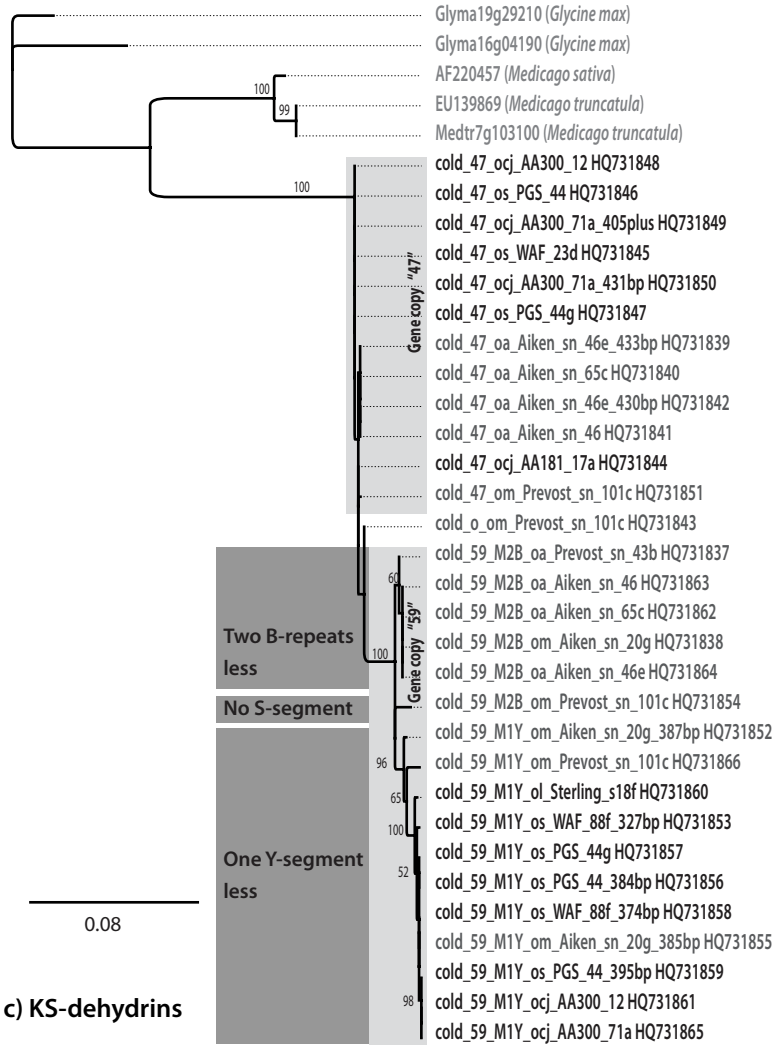
***Oxytropis* PR-10 genes (next page)**



a) Pathogenesis related class 10 (PR-10) (continued)



b) Ripening-related proteins



c) KS-dehydrins

Table 4.4 Estimation of copy number from sequence data for the genes and gene families surveyed in *Oxytropis* species[a] and number based on available data from public sequence databases for other Fabaceae.

Gene or gene family	Phytozome 5.0 legume Cluster	Number of gene copies for each species					
		Mt [b]	Gm [b]	Oa	Om	Ocj	Os
PR-10	22147878	1	10	3	3 [c]	3	4
Ripening-related proteins	22116173	7	6	3	4	3 [d]	4
KS-dehydrins	22130070	1	2	2	5 [d]	2	2
lhcaIII	22156497	1	4	1	1 [d]	1	1
lhcbI	22148038	3	2	1	1 [d]	1	1

[a] Number of *Oxytropis* gene copies was estimated from the number of different sequences per plantlet, their pairwise divergence (Supplementary Table S4.4), and the phylogenetic trees topology (Fig. 4.3).

[b] Counts for *Medicago* and *Glycine* report the Phytozome gene data.

[c] Due to the presence of many ambiguous peaks, some PR-10 sequences from *O. maydelliana* were excluded from analyses, but since they are similar to estimated gene copy "13" (data not shown), this copy is counted present in *O. maydelliana*.

[d] Ambiguous peaks in chromatograms from *O. maydelliana* for *lhcbI* (HQ731870 and HQ731871), *lhcaIII* (HQ731878) and KS-dehydrin (HQ731866, HQ731852, and HQ731855) suggest that actual number of genes could be higher.

Mt: *Medicago truncatula*; Gm: *Glycine max*; Oa: *Oxytropis arctobia*; Om: *O. maydelliana*; Ocj: *O. campestris* subsp. *johannensis*; Os: *O. splendens*.

4.5 Discussion

By sequencing genes from arctic and temperate *Oxytropis* species genomic DNA, followed by detection of codons under selection and phylogenetic analyses, we clarify the selective pressure acting on *Oxytropis* PR-10, ripening-related protein and KS-dehydrin gene families, and on light-harvesting proteins LHCAIII and LHCBI. These genes and gene families were previously shown to play a potential role in plant adaptation to the Arctic (Archambault and Strömvik 2011). Now we have shown that the gene families overexpressed in arctic species are characterized by negative selection maintained at a higher level in the arctic species. These genes were duplicated prior to speciation of arctic taxa.

4.5.1 The PR-10, ripening-related proteins and KS-dehydrins gene families conserved a modest size in *Oxytropis*

Pairwise sequence divergence, qPCR experiment and phylogenetic analyses were performed on *Oxytropis* gene families to answer one of our initial questions: is there evidence for more than one copy in the genome of the four *Oxytropis* spp. surveyed. Given the isolation methods that used specific primers pairs and the conservative logic, the count from sequence data is likely an underestimation, especially in the polyploid *Oxytropis* genomes. Indeed, copy number estimation from sequence data for PR-10 was lower than from qPCR fluorescence ratios. The latter suggested occurrence of likely more than four PR-10 genes in each *Oxytropis* genome in diploid species, and more than nine in the high polyploid. The number of PR-10 copies, which appears to vary among individuals of a species, is a feature not detected in other surveyed genes. In other species, there are ten PR-10 genes from *Glycine max* (Schmutz et al. 2010), eight from *Lupinus luteus* (Handschuh et al.

2007), thirteen from *Betula pendula* (Schenk et al. 2006), fourteen (nine functional) from *Vitis vinifera* (Lebel et al. 2010), 18 from *Malus domestica* (Gao et al. 2005), eight from *Prunus persica* and *P. dulcis* (Chen et al. 2008b) and eight from *Fragaria ananassa* (Musidlowska-Persson et al. 2007). The three to four different gene copies of ripening-related proteins discovered in *Oxytropis* also make a slightly lower count than in other Fabaceae (Hellsten et al. 2010; Schmutz et al. 2010), but more gene copies might still be discovered. The conservative estimation of two KS-dehydrins genes in the diploid and hexaploid *Oxytropis* is similar to *Medicago truncatula* and the diploidized tetraploid *Glycine max* gene counts (Hellsten et al. 2010; Schmutz et al. 2010). Other types of dehydrins occur in considerable number of copies in more distantly related genomes; nine in *Arabidopsis* (Bies-Etheve et al. 2008) and thirteen in *Hordeum* (Rorat 2006); but these dehydrins are not similar to the *Oxytropis* genes. By comparison to very large plant gene families composed of many hundreds of copies (Xu et al. 2009), the gene families studied here appear to have conserved a modest size over evolutionary time in Fabaceae species and possibly in the arctic and temperate *Oxytropis* as well. The sequence results also suggest that photosynthesis genes (*lhcaIII* and *lhcbI*) occur in low copy number in *Oxytropis*, as in other Fabaceae (Hellsten et al. 2010; Schmutz et al. 2010).

Although the precise genomic location of these PR-10, ripening-related proteins and KS-dehydrin is still unknown in *Oxytropis* genomes, these three gene families occur in one or a few blocks of tandem duplicated copies in the closely related diploid *M. truncatula* (Hellsten et al. 2010) and in *G. max* (Schmutz et al. 2010), and also in all other described plant genomes (Cannon et al. 2004; Gao et al. 2005; Bies-Etheve et al. 2008; Hundertmark and Hincha 2008; Lebel et al. 2010). They are therefore suspected to also occur in one or a few blocks of tandem gene copies in *Oxytropis* genomes. Among possible duplication mechanisms (Freeling 2009), ancestral whole genome or large segmental duplications, prior to an

Oxytropis-Medicago split, can be excluded, because they would induce a different tree topology, where the several clades observed would each contain genes from different related genera. Duplication by mobile elements (transposons) commonly produces copies devoid of introns (Freeling 2009), whereas all the *Oxytropis* genes surveyed possess as many intron as their relatives.

Theories in gene family evolution propose that genes retained (not silenced) after tandem duplications tend to belong to the "response to stimulus category" (Michelmore and Meyers 1998; Edger and Pires 2009; Freeling 2009), a category of gene that includes PR-10, ripening-related proteins and KS-dehydrin. Our data suggest that copy number in not different in arctic than in temperate diploid species. The count may be higher in the polyploid species, but only for some gene families. The occurrence of multiple copies prior to speciation may provide reservoirs for evolutionary innovations to allow closely related plant lineages to model these proteins sequence and structure precisely according to their needs.

4.5.2 PR-10 genes show a unique pattern of codons under selection

We investigated whether the sharp increase or decrease expression level previously discovered in response to stimulus genes in arctic plantlets (Archambault and Strömvik 2011) is accompanied by unique evolutive constraints in the arctic protein sequences compared to non-arctic ones. To this end, codons under positive and negative selection were detected from the gene copies isolated here from two arctic and two temperate *Oxytropis* species. The analyses included either all four species, arctic-only or temperate-only species.

These steps allowed us to answer a second question: 2) Do PR-10 and KS-dehydrin gene copies, found overexpressed in arctic species (Archambault and

Strömvik 2011), have higher levels of codon selection in the arctic species than in the temperate ones? Results clearly indicate that for genes of these two gene families, evolutive constraints are weaker in temperate species than in the arctic species. Another relevant result is that a slightly high number of codons under positive selection were identified in the complete set of *Oxytropis* PR-10 sequences and not when arctic and temperate *Oxytropis* PR-10 sequences are analyzed separately. Since positive selection indicates historical advantage of novel variants, this pattern in *Oxytropis* suggests that developing new PR-10 variants was required when plants evolved between contrasting environments. Recruiting novel protein variants in that situation may however not be common to all genes overexpressed in arctic species, because positive selection is nearly absent from the analyses of KS-dehydrin genes, which were also found overexpressed in the arctic *Oxytropis* (Archambault and Strömvik 2011).

The PR-10 and ripening-related proteins are two protein families weakly similar at the sequence level (Osmark et al. 1998), but with very similar three dimensional conformation. Both possess a typical internal cavity able to bind ligands such as the cytokinin plant hormones (Biesiadka et al. 2002; Mogensen et al. 2002; Koistinen et al. 2005; Fernandes et al. 2009; Lytle et al. 2009), and for a few proteins, brassinosteroids as well (Markovic-Housley et al. 2003; Koistinen et al. 2005; Lytle et al. 2009). Our previous discovery, that they have a contrasting expression pattern in arctic and temperate plantlets, was thus highly intriguing (Archambault and Strömvik 2011). Accordingly, the third specific question of this study asks 3) Do PR-10 and ripening-related proteins have similar or contrasting codon selection patterns in *Oxytropis*? The observation that there are almost no positively selected codons in the *Oxytropis* ripening-related proteins suggest that they may have an evolutionary history distinct from the PR-10, described above.

Minor sequence variants in specific protein regions have impact on the ligand-binding specificity (Pasternak et al. 2005; Fernandes et al. 2008), which warrants a more careful inspection of the codons under selection location. This inspection reveals that most positively selected codons in PR-10 are located in ß sheets and in the terminal variable region, which may therefore be responsible for the PR-10 novelties that evolved when colonizing stressful environments. An important role for PR-10 ß sheets represents a novel hypothesis, since the critical regions are usually considered to be the short-conserved glycine-rich loop (Biesiadka et al. 2002; Pasternak et al. 2006; Berkner et al. 2009; Lytle et al. 2009) and the last α helix itself with its upstream variable region (Pasternak et al. 2005; Fernandes et al. 2008). Collectively, these observations allow us to answer that the pattern of codons under selection of the PR-10, where novel protein variants were advantageous during evolution between contrasting environment, and where sequence evolution is then under increased negative selection in the harsher environment, is highly distinct and does not apply to the ripening-related proteins. The unique pattern of arctic PR-10 sequence is not solely due to common ancestry since these do not group into a clade. A future question is whether the cavity shape and by extension, the ligand-binding specificity, is more uniform for genes evolving under a higher negative selection (arctic PR-10) or for genes with fewer positively selected sites (ripening-related proteins).

It is difficult to interpret the ripening-related proteins selection pattern because their biological function is still uncertain (Lytle et al. 2009). The two main biological functions of PR-10 are, on the other hand, better described and relate to development by storing and carrying plant hormones (Srivastava et al. 2006), and to pathogenesis response by cleaving foreign RNA (Liu and Ekramoddoullah 2006). Because the Arctic is poor in endophytes, we previously suggested that arctic *Oxytropis* may not require exceptional pathogen resistance (Archambault and

Strömvik 2011). Arctic plants, however, do exhibit a characteristic growth form, such as the extremely compact stature of the cushion-like *Oxytropis arctobia* (Aiken et al. 2007), that could be mediated by a unique control in hormone sequestration and delivery in arctic species, possibly linked to the unique PR-10 sequence evolution in arctic species.

4.5.3 Selective pressure in *Oxytropis* cold dehydrins suggests a critical role for the uncharacterized Y-segment

The implication of dehydrin genes in cold response (Wong et al. 2006) and resistance (Rorat 2006; Remus-Borel et al. 2010) is well known. The many dehydrin roles may be mediated by a protective binding of the unstructured (Battaglia et al. 2008) random coil structure of the K-segment to lipid membranes (Rahman et al. 2010) and partially denatured proteins (Close 1996), or a general water binding (Rinne et al. 1999). Our discovery that *Oxytropis* dehydrins are of a novel type (K-like - Y_4 – K– S) limits extrapolation on their molecular mechanisms, but given their expression profile (Archambault and Strömvik 2011) they certainly are involved in arctic adaptation. We also show that the *Oxytropis* KS-dehydrins are under less stringent selection than the other genes surveyed. It should be noted here that codons under negative selection are detected when nonsynonymous substitution rates (d_N) are significantly lower than synonymous substitution rates (d_S). Amino acid sites may therefore be conserved between sequences, but not under negative selection if d_N is not significantly lower than d_S, as in the *Oxytropis* K-segment and S-segment. This finding is consistent with the low similarity we found with dehydrins from other species and with their biological roles in binding other molecules that can be achieved without a specific structure (Mouillon et al. 2008). For instance, two

KS-dehydrins that differ in their K-segment arrangement and precise sequence still similarly maintain their unstructured state (Mouillon et al. 2008). Dehydrins are likely able to achieve their biological function by maintaining similar, but not the same amino acid at particular codons positions.

Still, within the freely evolving set of *Oxytropis* KS-dehydrins, two regions show codons under negative selection: the S-segment, and the repeats that resemble Y-segment. The S-segment may be phosphorylated to facilitate binding of calcium ions and may act as nuclear localization signal (Battaglia et al. 2008) to various subcellular localization (cytoplasm, nucleus, plasma membrane, mitochondria, vacuole, and endoplasmic reticulum). Although the Y-segment presence or absence distinguishes different dehydrin subgroups (Battaglia et al. 2008), a functional role for this region is currently unknown. We show here that the glutamate and proline codons in most repeats are under negative selection in *Oxytropis* in the arctic sequences. Because of their physical properties, these amino acids are often conserved in protein evolution, but the negative selection pressure may also reflect a critical function, especially useful in the arctic, that has yet to be elucidated. The discovery of more stringent evolutive constraints in the arctic KS-dehydrins suggest their sequence fits tightly the requirements for a successful colonization of the rigorous arctic habitats.

Overall, our results demonstrated that the three gene families surveyed, PR-10, ripening-related proteins and KS-dehydrins were all duplicated prior to the *Oxytropis* genus diversification. Although both PR-10 and KS-dehydrin are expressed in arctic species, only PR-10 show evidence for elevated negative constraints in the arctic species. The few stringent evolutive constraints detected in the arctic KS-dehydrins cluster in the Y-segment. Our results also suggest that evolving novel protein variants in PR-10 genes was required for initial adaptation to the arctic environment. Taken together these results clarify the evolutionary

processes in action in arctic *Oxytropis* for genes that were suspected to have important biological role but remained to be characterized in these plants. Among possible future directions, studying the duplication tempo and confirming gene location within genomes in additional arctic and temperate *Oxytropis* species, and in the sister genus *Astragalus* would confirm whether tandem duplications in stimulus response genes are a prerequisite or a short-term consequence of Arctic colonization.

4.6 Acknowledgements

This work was supported by a grant from Natural Sciences and Engineering Research Council of Canada (NSERC) to MVS and by the Centre Sève (Le Fonds québécois de la recherche sur la nature et les technologies, FQRNT). AA was supported by a scholarship from FQRNT, and by Eben-Hopson and Girardin-Vaillancourt fellowships. The authors would like to thank Dr. Susan Aiken, Dr. Danielle Prévost, Dr. Carolyn Parker, Dr. Tracy Sterling and Plant Gene Resources of Canada (Agriculture and Agri-Food Canada, Saskatoon Research Centre) for providing *Oxytropis* seeds; Dr. Simon Joly for helpful comments on the manuscript, and Dr. Mohamed Hijri for access qPCR instrument.

Chapter 5

5 Connecting text to Chapter 5

Transcriptomes from two arctic *Oxytropis* species (*O. arctobia* and *O. maydelliana*) share some characteristics when compared to transcriptome of temperate species of the same genus (Chapter 3). However, the accepted taxonomic classification for these species, placed respectively in the *Arctobia* and the *Orobia* section of the genus, was our best estimation of their evolutionary relationships. One objective of Chapter 5 is to provide a picture of the evolutionary relationships in the *Oxytropis* genus gained from nuclear sequence data. This knowledge could rule out a long evolutionary common history of arctic species as the explanatory variable for similarity in expression profiles, rather than relatively independent but similar strategies for long-term colonization of the Arctic.

To this end, the nuclear ribosomal ITS (Internal Transcribed Spacer) region are isolated and sequenced from many individuals of species previously analyzed at the transcriptome level (Chapter 3), complemented by many others from public sequence databases. Phylogenetic analyses and statistical parsimony network method are applied to the sequence alignment that includes 9 arctic species out of a total of 30, and that covers a broad range of the *Oxytropis* diversity. Results confirm that the two arctic *Oxytropis* previously analyzed (Chapter 3 and Chapter 4) are not closely related, and further clarifies that the different arctic species evolve from many different temperate ancestral lineages, therefore excluding an adaptive radiation in arctic *Oxytropis*.

This chapter has been published in August 2012 in Botany (Archambault and Strömvik 2012a) DOI:10.1139/b2012-023; with the following keywords: *Oxytropis*, arctic, Internal Transcribed Spacer ITS, sequence evolutionary analyses, independent origin. It is included here with permission from NRC Research Press,

under the license number 3006530369165. Nucleotide sequence data reported are available in the DDBJ/EMBL/GenBank databases under the accession numbers HQ176472 to HQ176487.

5.1 Evolutionary relationships in *Oxytropis* species, as estimated from the nuclear ribosomal internal transcribed spacer (ITS) sequences point to multiple expansions into the arctic

5.1.1 Abstract

Species of the *Oxytropis* genus are distributed in the Northern hemisphere, especially in alpine and arctic areas. Although comprehensive taxonomic treatments exist for local floras, an understanding of the evolutionary relationships is lacking for the genus as a whole. To determine if different ancestral *Oxytropis* species colonized the North American Arctic separately, as suggested by taxonomy, we sequenced the nuclear ribosomal internal transcribed spacer (ITS) region from 16 *Oxytropis* specimens, including four species that were used in a previous transcriptome study. In addition, 81 other *Oxytropis* ITS sequences were retrieved from public sequence databases and included in the analysis. The whole dataset was analyzed using phylogenetic trees and statistical parsimony networks. Results show that all *Oxytropis* ITS sequences are very similar. Furthermore, at least six lineages evolved from different temperate ancestors to colonize the North American Arctic. This pattern is believed to be typical of the arctic flora. Additionally, the sequence relationships analyses confirm that the *Phacoxytropis* subgenus may be ancestral in *Oxytropis*.

5.1.2 Résumé

Les espèces du genre *Oxytropis* sont distribuées dans l'hémisphère Nord, surtout dans les régions alpines. Bien que des traitements taxonomiques exhaustifs existent pour les flores locales, il n’existe pas encore de compréhension globale des relations évolutives pour le genre dans son ensemble. Pour déterminer si, tel que le

suggère la taxonomie, différentes lignées d'*Oxytropis* auraient colonisées séparément l'Arctique nord-américain, nous avons séquencé l'ADN de la région nucléaire ribosomale ITS pour 16 spécimens d'*Oxytropis*, dont quatre utilisées dans une étude transcriptomique précédente. En outre, 81 autres séquences d'*Oxytropis* de bases de séquences publiques ont été incluses dans l'analyse. L'ensemble de données a été analysé à l'aide d'arbres phylogénétique et de réseau de parcimonie statistique (TCS). Les résultats montrent que toutes les séquences d'*Oxytropis* sont très similaires. Par ailleurs, au moins six lignées ont évolué à partir de différents ancêtres tempérés pour coloniser l'Arctique nord-américain. Ce modèle est typique de la flore arctique. De plus, les analyses de séquences confirment que la des membres du sous-genre *Phacoxytropis* pourraient être le groupe ancestral.

5.2 Introduction

Although rarely dominant, arctic *Oxytropis* are ecologically important since they are grazed by herbivores (Mulder and Harmsen 1995; Gauthier et al. 1996). They also develop nodules by symbiosis with specific rhizobia (Prevost et al. 1987), that may improve nitrogen content in an otherwise poor arctic soil (Russell et al. 1940). The genus of 310 (Langran et al. 2010) to 450 (Malyshev 2008) species includes up to six subgenera and 25 sections (Malyshev 2008). Many *Oxytropis* species are alpine, and the center of diversity is Central Asia with 153 to 166 species (Malyshev 2008). North America (from Mexico to the Canadian Arctic) is home to 22 (Welsh 2001) to 24 (Malyshev 2009) *Oxytropis* species. The whole Arctic harbors between 22 to 45 species according the PanArctic Flora (Elven 2011). The range in species number is due to the difference in taxonomic treatments, where lower ranking taxa are either merged with other species, or split as a separate species.

Comprehensive taxonomic treatments exist for Russian (Yakovlev et al. 1996), Chinese (Langran et al. 2010), North American (Welsh 2001) and arctic (Elven 2007) floras. However, taxonomists typically analyzed specimens from local geographical areas and the few molecular studies on *Oxytropis* also addressed local questions about distinction of related or rare taxa using highly variable markers (Jorgensen et al. 2003; Artyukova et al. 2004; Chung et al. 2004; Schonswetter et al. 2004). A broad overview of relationships between *Oxytropis* species from different continents is not yet available. This is especially problematic for arctic species since lineages from different continents came into close contact after northward migration into the area. Classification of arctic species that are found in five taxonomic sections within two subgenera (Elven 2011) suggests that five *Oxytropis* lineages expanded to the Arctic, but this has yet to be confirmed by independent molecular markers.

The basic chromosome number for *Oxytropis* is 8 (2n = 2x = 16), but ploidy level differs among species (Ledingham 1957; Ledingham 1960; Ledingham and Rever 1963). *O. maydelliana*, for example, has 96 chromosomes (Elven 2011). The scarcity of studies on *Oxytropis* is in sharp contrast to the intensive molecular phylogenetic efforts carried out in *Astragalus*, the *Oxytropis* sister genus, which includes more than 2500 species, and has contributed to the understanding of the general question of adaptive radiation (e.g. Wojciechowski et al. 1993; Wojciechowski et al. 1999; Wojciechowski 2005).

In a recent study, we compared the subtracted transcriptomes of plantlets from four North American species, *O. arctobia* (arctic diploid), *O. maydelliana* (arctic polyploid), *O. campestris* subsp. *johannensis* (temperate polyploid) and *O. splendens* (temperate diploid) to identify potential molecular mechanisms for arctic adaptation (Archambault and Strömvik 2011). The sequences of several identified genes, such as PR-10, ripening-related proteins, KS-dehydrins and light harvesting

proteins, were further analyzed for potential selective pressure (Archambault and Strömvik 2012b).

There is a need to confirm, using independent markers, the relationships suggested by taxonomic classification of arctic *Oxytropis* species. Phylogenetic sequence analyses are standard methods to infer relationships among organisms, and a phylogenetic tree including arctic and temperate *Oxytropis* lineages would further allow the relative timing of arctic lineages evolution to be estimated. This knowledge may suggest basic evolutive trends in *Oxytropis*, such as evolution of ploidy levels, of plant traits or of geographic distribution. Sequence network analyses are, on the other hand, useful for taking into account a weak signal in a visually informative manner (Huson and Bryant 2006).

In the present study we provide an analysis of the general evolutionary relationships within the *Oxytropis* genus, with an emphasis on the North American arctic species. Our main goal was to determine the *Oxytropis* lineages that expanded to the North American Arctic. Nuclear ribosomal ITS regions from *Oxytropis* and *Astragalus* samples were isolated in our laboratory or downloaded from public sequence databases. A phylogenetic tree and a statistical parsimony network are presented, the many lineages involved in North American Arctic colonization are identified and taxonomic implications for the *Oxytropis* genus are discussed.

5.3 Material and Methods

5.3.1 Plant material and PCR amplification protocol

Seeds were scarified by rubbing with sand, sterilized in 70% ethanol for 3 min then in 2% sodium hypochlorite for 10 min, followed by eight rinses in distilled

autoclaved water, and placed on 1/2 Murashige and Skoog Basal Medium with Gamborg's vitamins (catalog number M0404 Sigma, Oakville, Ontario) agar plates. The plated seeds were stratified at 4°C in darkness for 20 days, in order to stimulate and synchronize germination. The seed sources included the following *Oxytropis* species: *O. campestris* (L.) DC subsp. *johannensis* (Fernald) Blondeau & Gervais from Quebec, *O. lambertii* Pursh from New Mexico, *O. splendens* and *O. sericea* Nutt. from Alberta and Saskatchewan, *O. podocarpa* A. Gray, *O. arctobia* Bunge, *O. maydelliana* Trautv and *O. deflexa* (Pall.) DC. subsp. *foliolosa* (Hook.) Cody from Baffin Island (Supplementary Table S5.1). Specimens for some of the material collected by Annie Archambault are deposited in the McGill University, Macdonald Campus Herbarium (MTMG).

Genomic DNA was extracted using DNeasy kits (Qiagen, catalog number 69104) from whole plantlets grown in Petri plates. The nuclear ITS was amplified by standard PCR using primers AB101 and AB102 (Sun et al. 1994). The reaction was carried out in a 50 μl volume with 1X Buffer (Invitrogen catalog number 18038), and with a final concentration of 2 mM $MgCl_2$, 0.2 mM (each) dNTP, 0.4 μM each primer, 0.05 unit Taq DNA polymerase per μl (Invitrogen catalog number 18038) and a total of 40 μg of genomic DNA (or less when quantity was limited). The following cycling program was applied: initial denaturation at 96 °C for 2 minutes, 30 cycles consisting of [96 °C for 15 sec, 55 °C for 30 sec, 72 °C for 3 min], followed by a final extension at 72 °C for 10 min. A small aliquot of each amplification was analyzed by gel electrophoresis to confirm amplified fragment length. The PCR product was directly sent for purification and sequencing from both ends using the AB101 and AB102 primers (Sanger sequencing on a 3730xl DNA Analyzer of Applied Biosystems at the McGill University and Génome Québec Innovation Center). The two sequences from each fragment were assembled using the assembly tool in the Geneious software suite (Drummond et al. 2008) and chromatograms

were reviewed and edited. All newly isolated sequences are submitted to GenBank (accessions no. HQ176472 to HQ176487).

5.3.2 Sequence analysis

An *O. arctobia* ITS sequence was used as blastn query sequence to retrieve (in May 2010) additional *Oxytropis* ITS sequences available from NCBI GenBank (Sayers et al. 2010). Only sequences with fewer than three ambiguous sites (N, R, Y, S, W, K, M for instance) were retained for further analyses (Supplementary Table S5.2), and the few sequences of the ITS1 and the ITS2 sections that were from different GenBank entries, despites coming from a single plant accession, were concatenated. Sequences were aligned using the "Geneious alignment" tool in the Geneious software suite (Drummond et al. 2008) and the alignment was manually corrected. The AB101 and AB102 primers are located further upstream in the 5.8S and further downstream in the 28S conserved genes than the less robust ITS4 and ITS5 primers commonly used. Amplifications with AB101 and AB102 primers therefore generate longer sequences than the typical ITS sequences of other phylogenetic studies, and in the final *Oxytropis* alignment, the protruding ends from our reactions were trimmed to facilitate comparison. The final *Oxytropis* dataset included 97 individuals, from 30 *Oxytropis* species (Supplementary Table S5.2).

The ITS sequences longer than 500 bp from *Astragalus*, the sister genus of *Oxytropis* (Lavin et al. 2005) were also downloaded from NCBI GenBank and aligned using Geneious. Two series of phylogenetic analyses were then performed to determine the ancestral *Oxytropis* lineages. The 97 *Oxytropis* sequences were first aligned with the 422 *Astragalus* sequences (from 340 species or subspecies) and a phylogenetic analysis was performed with PHYML (Guindon and Gascuel 2003)

using Geneious, with the sequence evolution model determined by MrModelTest (Nylander 2004). Both the tree and the alignment were deposited in TreeBASE (TB2: S12147). According to this large phylogenetic tree, the group of 25 *Astragalus* sequences most closely related to *Oxytropis* were identified, and the second analysis step included only those 25 as outgroup with the 97 *Oxytropis* sequences as ingroup. To estimate the lineages that contributed to arctic *Oxytropis*, phylogenetic analyses (TB2:S12147) were performed with PHYML and MrBayes (Huelsenbeck and Ronquist 2001) implemented in Geneious, with the sequence evolution model determined by MrModelTest (Nylander 2004). Ancestral states were reconstructed for the continental geographic origin of the specimen (Arctic, North America, Asia, Europe), using Mesquite (Maddison and Maddison 2011), with the "Trace Character History" module applying the parsimony reconstruction method on the Maximum Likelihood tree derived from the PHYML analysis. The ploidy level, from the Index of Plant Chromosome Number (IPCN) available in Tropicos website (Goldblatt 2007) or indicated in the Panarctic Flora (Elven 2011), was added after terminal nodes on the phylogenetic tree.

The level of divergence within the 97 *Oxytropis* and the 422 *Astragalus* sequences was compared using the Species Delimitation tool in Geneious (Masters et al. 2011). Two measures of average pairwise tree distance are calculated with this tool: Intra-group distances (Intra-Dist) and Inter-group distances (Inter-Dist).

Statistical parsimony implemented in TCS 1.21 (Clement et al. 2000) and NeighborNet in SplitsTree4 (Huson and Bryant 2006) networks were also generated. For statistical parsimony (TCS), sequences with ambiguous sites were placed towards the bottom of the sequence alignment (Joly et al. 2007), gaps were treated as missing characters and the connection limit was set to 95%. Therefore, although no additional statistic is calculated on a TCS network, an edge corresponds to the probability of parsimony (here 95%) calculated by DNA pairwise differences.

The split network was calculated from uncorrected distances from DNA sequences and averaged ambiguous characters. Splits were calculated with the NeighborNet algorithm, which computes a set of incompatible splits to represent incompatible and ambiguous signals in a dataset; and an ordinary least square variance network was built with the equal angle option. Support on edges was also assessed with bootstrap resampling of 1000 replicates (Huson and Bryant 2006), and high support is indicated by a thick line. The sequences located internally in the network were identified since they are inferred to be ancestral (Huson and Bryant 2006). An attempt to date nodes with the BEAST program (Drummond and Rambaut 2007) was made using the only internal dated node available (Wojciechowski 2005).

5.4 Results

5.4.1 *Oxytropis* species from diverse sections of the genus show little diversity, compared to *Astragalus*

The sequence dataset was assembled to address evolutionary question on the North American arctic species of the *Oxytropis* genus. The final dataset for the phylogenetic analyses consisted of sequences of the nuclear ribosomal internal transcribed spacer (ITS) loci from 97 specimens of 30 *Oxytropis* species (Table 5.1), spanning most of the genus diversity (Malyshev 2008), and 422 sequences from the sister genus *Astragalus*. Most sequences were downloaded from GenBank, but sixteen were obtained by de novo sequencing (Supplementary Table S5.2).

The 97 ITS sequences from *Oxytropis* accessions (Supplementary Table S5.2) were aligned in a 715 base long matrix, which included 57 variable sites, and three insertion sites of one and two bp. The matrix length became 763 bp long when the 422 *Astragalus* sequences were included in the alignment, which comprised many

indels, of maximum length of 3 bp. The first results gained from a phylogenetic analysis on these 519 sequences (data not shown, but available from TreeBASE TB2:S12147) were measures of average pairwise tree distances, which use the tree as data, rather than the underlying sequence alignment. The measure of intra-group distances (Intra-Dist) is almost five times lower within *Oxytropis* than within *Astragalus* (Table 5.2), although these two datasets contain a similarly small fraction of their constituent species (approximately 10% for *Oxytropis* and 17% for *Astragalus*). This difference indicates that species of the *Astragalus* genus are more diverse than those of *Oxytropis*. To confirm that this difference in intra-group distances was not inflated by the high number of *Astragalus* sequences included, the average pairwise tree distances were also calculated on the reduced dataset that includes only 25 *Astragalus* sequences. In this tree, the difference in the measure is similar and even slightly higher than with the tree that includes 422 *Astragalus* sequences. The intra-group distances is also six times lower than the average pairwise tree distance of the *Oxytropis* to the *Astragalus* group, indicating that the two genera are quite distinct.

Table 5.1 Overview of agreement and disagreement between sequence grouping and taxonomic classification of *Oxytropis* species into sections and subgenera (Langran et al. 2010; Malyshev 2008).

Subgenus	Section	Species included in the sequence analyses	Informal groups [c]
Oxytropis	*Arctobia*	*O. arctobia, O. nigrescens*	-
		O. podocarpa	-
	Gloeocephala	*O. borealis* var. *viscida*	Eta
	Orobia	*O. campestris, O. sericea, O. campestris* subsp. *johannensis, O. arctica* var. *arctica*	
		O. maydelliana, O. arctica var. *barnebyana, O. arctica* var. *koyokukensis*	Delta
		O. besseyi var. *ventosa*	-
		O. campestris subsp. *gracilis, O. arctica* var. *barnebyana*	Gamma
		O. arctica var. *arctica, O. lambertii*	Zeta
	Verticillares (syn. *Baicalia*) [a]	*O. splendens, O. verticillaris, O. ochrantha, O. racemosa, O. verticillaris, O. splendens, O. bicolor*	Epsilon
		O. oxyphylla, O. ischanica, O. chankaensis [b]	-
	Xerobia	*O. anertii, O. ciliata*	-
	Oxytropis	*O. grandiflora*	-
	Eumorpha	*O. caerulea,*	-
		O. filiformis	Zeta
	Leucopodia	*O. squammulosa,*	Beta
Tragacanthoxytropis	*Lycotriche*	*O. aciphylla*	
Phacoxytropis	*Mesogaea*	*O. pilosa, O. glabra*	
		O. deflexa subsp. *foliolosa, O. deflexa*	Alpha
	Janthina	*O. oreophila*	Zeta
Physoxytropis	*Physoxytropis*	*O. multiceps*	-

Note: -, Indicates species not assigned to an informal group

[a] The section *Verticillares* (Malyshev 2009), corresponds to section *Baicalia* (Langran et al. 2010).

[b] In the Flora of China (Langran et al. 2010), *O. chankaensis* is synonymous to *O. oxyphylla*.

[c] Considering the phylogenetic tree (Fig. 5.1) and the sequence network (Fig. 5.2), congruent informal groups were recognized and named from Alpha to Eta.

Table 5.2 Average pairwise tree distance within species of the *Oxytropis* and within species of the *Astragalus* genera; as well as between the two genera.

Distance measure	*Oxytropis* genus	*Astragalus* genus
Intra Dist[a]	0.011	0.051
Inter Dist[b]	0.063	0.063
Intra/Inter[c]	0.18	0.81

All measures were calculated on a phylogenetic tree that includes 97 *Oxytropis* sequences and 422 *Astragalus* sequences.

[a] Intra-group Distances: gives an indication of the sequence diversity within a group.

[b] Inter-group Distances: gives an indication on the degree of distinctiveness between *Oxytropis* and *Astragalus.*

[c] Ratio of the within-group genetic differentiation (Intra-Dist) to the distance to the nearest group (Inter-Dist).

5.4.2 Nuclear genomic sequences generally agree with traditional taxonomy, with some exceptions

Evolutionary relationships of *Oxytropis* samples were estimated by phylogenetic reconstructions of sequences of the nuclear ITS loci, using a restricted set of *Astragalus* sequences as outgroup (25 sequences) and by two network methods, to determine whether multiple *Oxytropis* lineages expanded to the Arctic. Monophyly of *Oxytropis* is confirmed in the phylogenetic tree (Fig. 5.1) by a branch supported with 100% Bayesian consensus. Relationships among *Oxytropis* species are by contrast poorly resolved and only fifteen branches receive support from the Bayesian consensus. Statistical Parsimony network (TCS) was also generated as the number of substitutions available within *Oxytropis* is small and the phylogenetic tree was poorly resolved. The attempt to date nodes with the BEAST program (Drummond and Rambaut 2007) was impracticable, likely due to the combined effect of disposing of only a few substitutions and only of one internal dated node for calculation (Wojciechowski 2005).

Considering both the phylogenetic tree (Fig. 5.1) and the statistical parsimony network (Fig. 5.2), seven informal groups (named Alpha to Eta) are designed to recognize congruent groups of sequences among the analysis methods (Table 5.1). The two network methods give a similar estimate of sequence relationships, and only the statistical parsimony network is further described (Fig. 5.2), to avoid redundancy, although the NeighborNet is available (Supplementary Fig. S5.1).

The Alpha group (*O. deflexa* and *O. amethystea*) is basal in the phylogenetic tree (Fig. 5.1), which suggests that it is ancestral, but it is at one tip of the sequence network (Fig. 5.2), which suggests the opposite. The Beta group is an assemblage of non-arctic species from the *Phacoxytropis*, the *Oxytropis* and the *Tragancoxytropis* subgenera (Table 5.1), that comes next to the root in the tree (Fig. 5.1), and in central-lateral position in the network (Fig. 5.2).

The majority of *Oxytropis* species sampled are included in a large clade supported by Bayesian consensus in the phylogenetic tree (Fig. 5.1). The Gamma group is found within this clade (Fig. 5.1), and is at one extremity of the network (Fig. 5.2). It is composed of the arctic members of the *Orobia* section. Similarly, the Delta group is composed of other arctic species of the *Orobia* section (Fig. 5.1). It is also found within this large clade (Fig. 5.1), and it connects to a central node of the network (Fig. 5.2). The Epsilon group is composed of five species of the *Verticillares* section and connects to three other nodes in the network (Fig. 5.2). The Zeta group is an assemblage of arctic (*O. arctica* var. *arctica*) and temperate samples from diverse geographical origins (Asia and North America) and from different sections of the *Oxytropis* subgenus (*Orobia, Verticillares* and *Eumorpha*) and it also includes one sample of the *Phacoxytropis* subgenus. The Zeta group is near central and directly connects with the Eta group in the network (Fig. 5.2), while it embeds the Eta group in the phylogenetic tree (Fig. 5.1). Eta group is uniquely composed of species of the *Orobia* and the *Gloeocephala* sections, both of the *Oxytropis* subgenus, and is located at one extremity of the network (Fig. 5.2).

Generally, sequences from the same *Oxytropis* species or subspecies are at the same or adjacent nodes in the statistical parsimony network (Fig. 5.2), and they are not separated by highly supported branches in the phylogenetic tree (Fig. 5.1). There are two exceptions, where specimens are more related to other species rather than their conspecific, in both analysis methods. The first case is the different samples of the Alaskan *O. arctica* subspecies and varieties where the *O. arctica* var. *arctica* samples are related to temperate species, while the *O. arctica* var. *barnebyana* and *O. arctica* var. *koyokukensis* samples are related to *O. maydelliana*, another arctic taxon. The second case is the Asian *O. aciphylla* of the *Tragancoxytropis* subgenus, where specimens are kept apart by two highly supported branches in the tree (Fig. 5.1) and in the network they are split in two groups not

directly connected (Fig. 5.2). The two closely related specimens (GQ422810 and GQ422807) are close geographically.

At the subgeneric taxonomic level, but not at the sectional level, both sequence analyses methods mostly agree with taxonomy (Table 1). All representatives of the massive *Oxytropis* subgenus are in a highly supported clade in the phylogenetic tree (Fig. 5.1) and are directly connected in the sequence network (Fig. 5.2). Within this subgenus, species from the same section (Table 1) are not particularly related (Fig. 5.1), nor found at adjacent nodes (Fig. 5.2), especially for the large *Orobia* and for the *Verticillares* sections. Notable disagreements between taxonomy and sequence grouping (Table 1) occur for the *Phacoxytropis* subgenus and for the *Arctobia* section. The single sequence representative of the *Janthina* section (*O. oreophila*) is very distantly located from other members of the *Mesogaea* section (Fig. 5.1 and Fig. 5.2), whereas they are both of the *Phacoxytropis* subgenus. The two *O. podocarpa* samples are distantly located from *O. arctobia* and *O. nigrescens* (Fig. 5.1 and Fig. 5.2), the other species of the *Arctobia* section.

Although ploidy level is still not determined for most of the *Oxytropis* species (Fig. 5.1), a few patterns are discernable. All basal species with known ploidy are diploids, while all species of the Delta group, which is not basal, are all higher polyploids.

Figure 5.1 Likelihood phylogenetic reconstruction (PHYML) of the nuclear ribosomal internal transcribed spacer (ITS) sequences for 30 species (97 sequences) within the *Oxytropis* genus, using *Astragalus* representatives as outgroup.

Characters states for the geographical origin, ploidy level and subgeneric classification (Malyshev 2008; Langran et al. 2010) are indicated by grey shaded boxes before the sequence name. Branch color refers to geographical origin, following the same legend as for the shaded boxes. Numbers above branches report Bayesian support, in the form of percentage of trees from Bayesian analysis with this branch. Scale bar denote number of substitution per nucleotide according to the PHYML analysis. The two or three letters code refers to the taxonomic section of the species Ar: *Arctobia*, Eu: *Eumorpha*, Gl: *Gloeocephala*, Ja: *Janthina*, Le: *Leucopodia*, Ly: *Lycotriche*, Me: *Mesogaea*, Or: *Orobia*; Ox: *Oxytropis*, Phy: *Physoxytropis*, Ve: *Verticillares*, Xe: *Xerobia*. Lineages leading to the arctic groups are identified with a snowflake symbol. Underlying light grey boxes with a bracket on the right indicate informal groups Alpha to Eta. Dating of one node from another study (Wojciechowski 2005) is reported here.

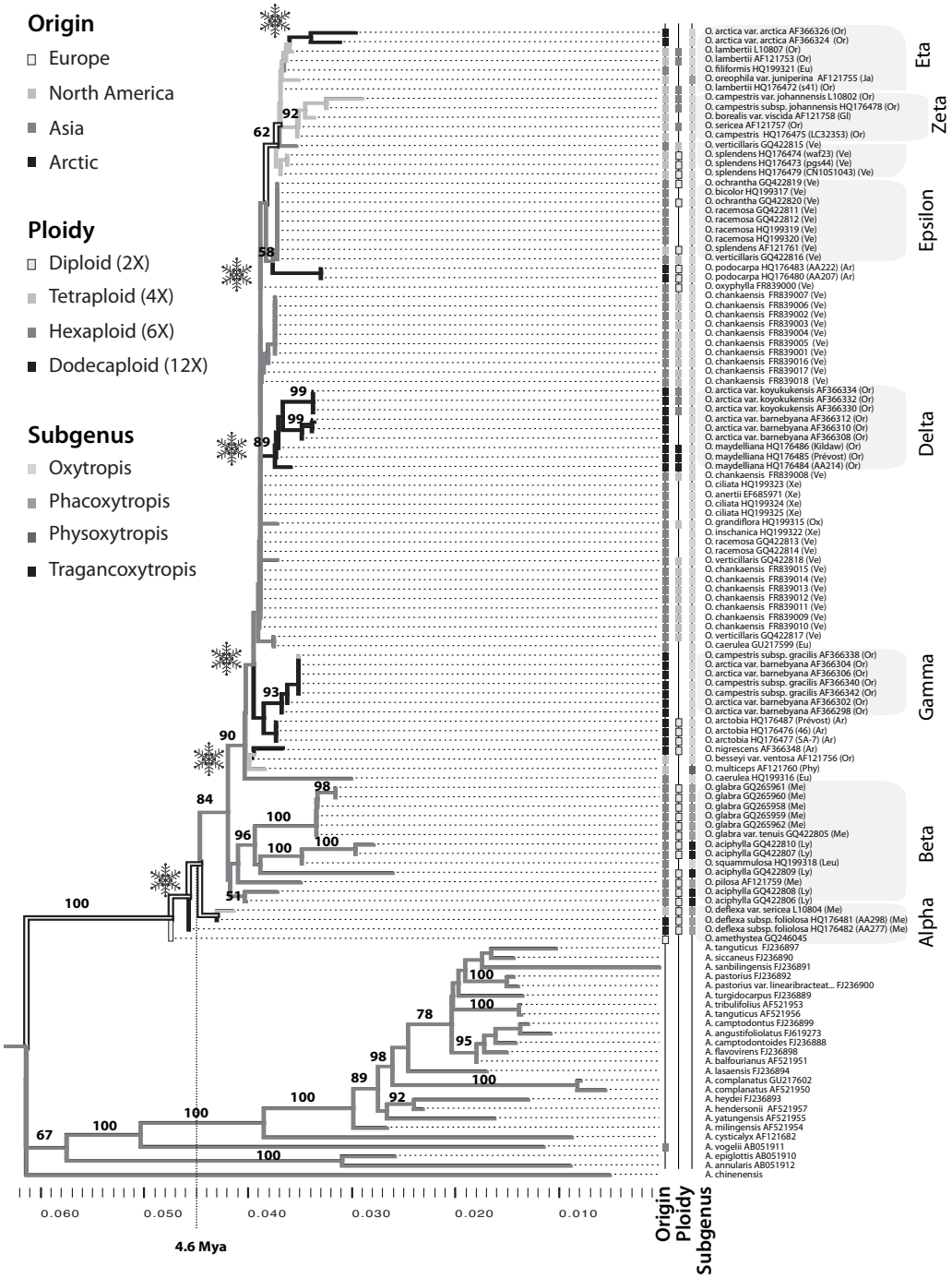
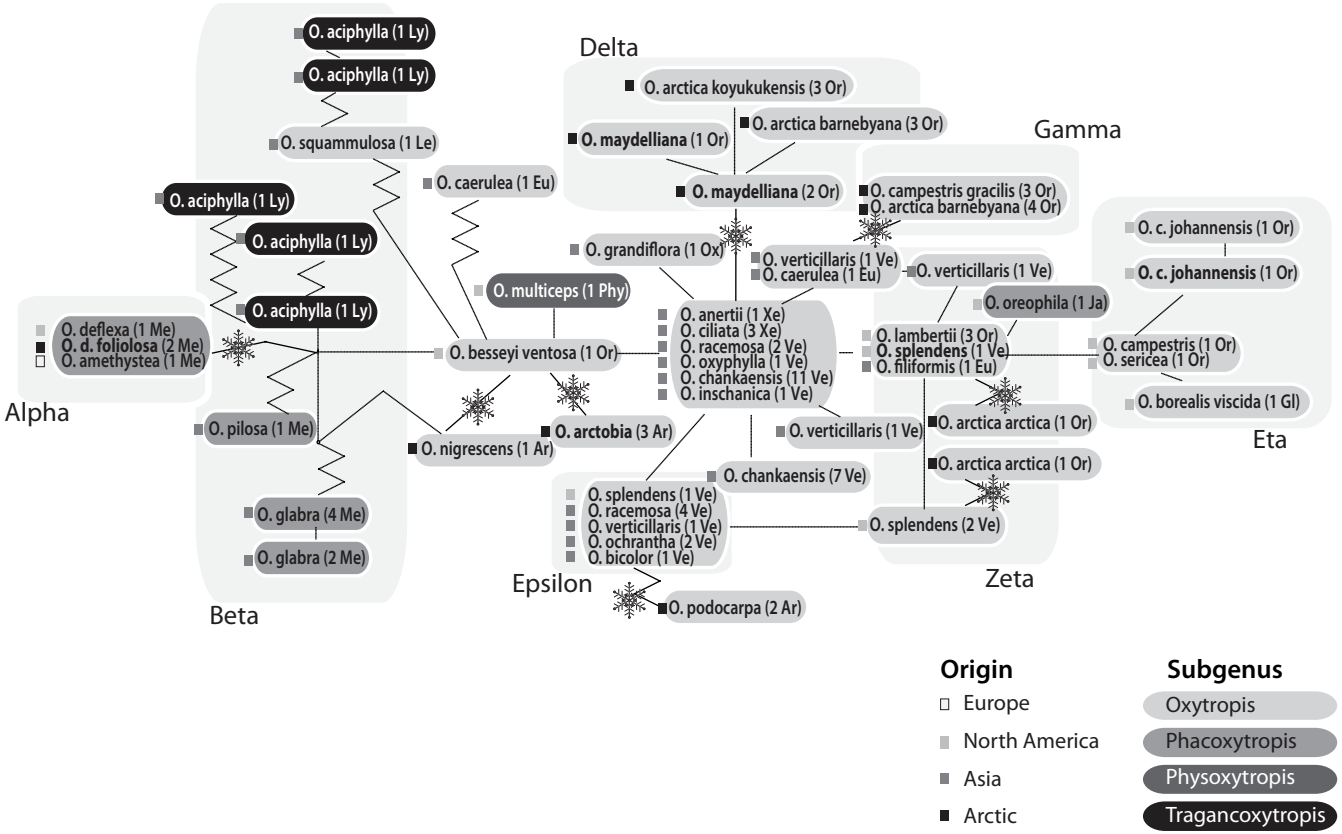

Origin
Europe
North America
Asia
Arctic
Ploidy
Diploid (2X)
Tetraploid (4X)
Hexaploid (6X)
Dodecaploid (12X)
Subgenus
Oxytropis
Phacoxytropis
Physoxytropis
Tragancoxytropis
O. arctica var. arctica AF366326 (Or)
O. arctica var. arctica AF366324 (Or)
O. lambertii L10807 (Or)
O. lambertii AF121753 (Or)
O. filiformis HQ199321 (Eu)
O. oreophila var. juniperina AF121755 (Ja)
O. lambertii HQ176472 (s41) (Or)
O. campestris var. johannensis L10802 (Or)
O. campestris subsp. johannensis HQ176478 (Or)
O. borealis var. viscida AF121758 (Gl)
O. sericea AF121757 (Or)
O. campestris HQ176475 (LC32353) (Or)
O. verticillaris GQ422815 (Ve)
O. splendens HQ176474 (waf23) (Ve)
O. splendens HQ176473 (pgs44) (Ve)
O. splendens HQ176479 (CN1051043) (Ve)
O. ochrantha GQ422819 (Ve)
O. bicolor HQ199317 (Ve)
O. ochrantha GQ422820 (Ve)
O. racemosa GQ422811 (Ve)
O. racemosa GQ422812 (Ve)
O. racemosa HQ199319 (Ve)
O. racemosa HQ199320 (Ve)
O. splendens AF121761 (Ve)
O. verticillaris GQ422816 (Ve)
O. podocarpa HQ176483 (AA222) (Ar)
O. podocarpa HQ176480 (AA207) (Ar)
O. oxyphylla FR839000 (Ve)
O. chankaensis FR839007 (Ve)
O. chankaensis FR839006 (Ve)
O. chankaensis FR839002 (Ve)
O. chankaensis FR839003 (Ve)
O. chankaensis FR839004 (Ve)
O. chankaensis FR839005 (Ve)
O. chankaensis FR839001 (Ve)
O. chankaensis FR839016 (Ve)
O. chankaensis FR839017 (Ve)
O. chankaensis FR839018 (Ve)
O. arctica var. koyukukensis AF366334 (Or)
O. arctica var. koyokukensis AF366332 (Or)
O. arctica var. koyokukensis AF366330 (Or)
O. arctica var. barnebyana AF366312 (Or)
O. arctica var. barnebyana AF366310 (Or)
O. arctica var. barnebyana AF366308 (Or)
O. maydelliana HQ176486 (Kildaw) (Or)
O. maydelliana HQ176485 (Prévost) (Or)
O. maydelliana HQ176484 (AA214) (Or)
O. chankaensis FR839008 (Ve)
O. ciliata HQ199323 (Xe)
O. anertii EF685971 (Xe)
O. ciliata HQ199324 (Xe)
O. ciliata HQ199325 (Xe)
O. grandiflora HQ199315 (Ox)
O. inschanica HQ199322 (Xe)
O. racemosa GQ422813 (Ve)
O. racemosa GQ422814 (Ve)
O. verticillaris GQ422818 (Ve)
O. chankaensis FR839015 (Ve)
O. chankaensis FR839014 (Ve)
O. chankaensis FR839013 (Ve)
O. chankaensis FR839012 (Ve)
O. chankaensis FR839011 (Ve)
O. chankaensis FR839009 (Ve)
O. chankaensis FR839010 (Ve)
O. verticillaris GQ422817 (Ve)
O. caerulea GU217599 (Eu)
O. campestris subsp. gracilis AF366338 (Or)
O. arctica var. barnebyana AF366304 (Or)
O. arctica var. barnebyana AF366306 (Or)
O. campestris subsp. gracilis AF366340 (Or)
O. campestris subsp. gracilis AF366342 (Or)
O. arctica var. barnebyana AF366302 (Or)
O. arctica var. barnebyana AF366298 (Or)
O. arctobia HQ176487 (Prévost) (Ar)
O. arctobia HQ176476 (46) (Ar)
O. arctobia HQ176477 (SA-7) (Ar)
O. nigrescens AF366348 (Ar)
O. besseyi var. ventosa AF121756 (Or)
O. multiceps AF121760 (Phy)
O. caerulea HQ199316 (Eu)
O. glabra GQ265961 (Me)
O. glabra GQ265960 (Me)
O. glabra GQ265958 (Me)
O. glabra GQ265959 (Me)
O. glabra GQ265962 (Me)
O. glabra var. tenuis GQ422805 (Me)
O. aciphylla GQ422810 (Ly)
O. aciphylla GQ422807 (Ly)
O. squammulosa HQ199318 (Leu)
O. aciphylla GQ422809 (Ly)
O. pilosa AF121759 (Me)
O. aciphylla GQ422808 (Ly)
O. aciphylla GQ422806 (Ly)
O. deflexa var. sericea L10804 (Me)
O. deflexa subsp. foliolosa HQ176481 (AA298) (Me)
O. deflexa subsp. foliolosa HQ176482 (AA277) (Me)
O. amethystea GQ246045
A. tanguticus FJ236897
A. siccaneus FJ236890
A. sanbilingensis FJ236891
A. pastorius FJ236892
A. pastorius var. linearibracteat... FJ236900
A. turgidocarpus FJ236889
A. tribulifolius AF521953
A. tanguticus AF521956
A. camptodontus FJ236899
A. angustifoliolatus FJ619273
A. camptodontoides FJ236888
A. flavovirens FJ236898
A. balfourianus AF521951
A. lasaensis FJ236894
A. complanatus GU217602
A. complanatus AF521950
A. heydei FJ236893
A. hendersonii AF521957
A. yatungensis AF521955
A. milingensis AF521954
A. cysticalyx AF121682
A. vogelii AB051911
A. epiglottis AB051910
A. annularis AB051912
A. chinenensis
Eta
Zeta
Epsilon
Delta
Gamma
Beta
Alpha
0.060
0.050
0.040
0.030
0.020
0.010
4.6 Mya
Origin
Ploidy
Subgenus

Figure 5.2 Evolutionary relationships among 30 *Oxytropis* species (97 sequences) calculated from the nuclear ribosomal internal transcribed spacer (ITS) sequences, shown as a statistical parsimony sequence network.

Each edge of the network represents a single mutation (Clement et al. 2000). Numbers in parenthesis after a species name indicate the numbers of individual harbouring this sequence. The two or three letters code refers to the taxonomic section of the species Ar: *Arctobia*, Eu: *Eumorpha*, Gl: *Gloeocephala*, Ja: *Janthina*, Le: *Leucopodia*, Ly: *Lycotriche*, Me: *Mesogaea*, Or: *Orobia*; Ox: *Oxytropis*, Phy: *Physoxytropis*, Ve: *Verticillares*, Xe: *Xerobia*. Lineages leading to the arctic groups are identified with a snowflake symbol. Subgenera of each species, as well as geographical origin are indicated by different grey shades, as seen in the legend. Underlying light grey boxes and brackets indicate informal groups Alpha to Eta. The *O. arctobia, O. maydelliana, O. splendens, O.campestris* subsp. *johannensis* analysed earlier (Archambault and Strömvik 2011) are marked in bold.


Alpha
O. deflexa (1 Me)
O. d. foliolosa (2 Me)
O. amethystea (1 Me)
Beta
O. aciphylla (1 Ly)
O. aciphylla (1 Ly)
O. squammulosa (1 Le)
O. aciphylla (1 Ly)
O. aciphylla (1 Ly)
O. aciphylla (1 Ly)
O. pilosa (1 Me)
O. glabra (4 Me)
O. glabra (2 Me)
O. caerulea (1 Eu)
O. multiceps (1 Phy)
O. besseyi ventosa (1 Or)
O. nigrescens (1 Ar)
O. arctobia (3 Ar)
Delta
O. arctica koyukukensis (3 Or)
O. arctica barnebyana (3 Or)
O. maydelliana (1 Or)
O. maydelliana (2 Or)
O. grandiflora (1 Ox)
O. verticillaris (1 Ve)
O. caerulea (1 Eu)
O. anertii (1 Xe)
O. ciliata (3 Xe)
O. racemosa (2 Ve)
O. oxyphylla (1 Ve)
O. chankaensis (11 Ve)
O. inschanica (1 Ve)
O. verticillaris (1 Ve)
O. chankaensis (7 Ve)
Epsilon
O. splendens (1 Ve)
O. racemosa (4 Ve)
O. verticillaris (1 Ve)
O. ochrantha (2 Ve)
O. bicolor (1 Ve)
O. podocarpa (2 Ar)
Gamma
O. campestris gracilis (3 Or)
O. arctica barnebyana (4 Or)
O. verticillaris (1 Ve)
O. oreophila (1 Ja)
O. lambertii (3 Or)
O. splendens (1 Ve)
O. filiformis (1 Eu)
O. arctica arctica (1 Or)
O. arctica arctica (1 Or)
O. splendens (2 Ve)
Zeta
O. c. johannensis (1 Or)
O. c. johannensis (1 Or)
O. campestris (1 Or)
O. sericea (1 Or)
O. borealis viscida (1 Gl)
Eta
Origin
Europe
North America
Asia
Arctic
Subgenus
Oxytropis
Phacoxytropis
Physoxytropis
Tragancoxytropis

5.4.3 Arctic species originate from temperate groups

Because the 26 sequences from nine different arctic taxa are found in different informal groups (Alpha, Delta, Gamma and Zeta) that also include temperate species (except for Gamma and Delta), a single adaptive radiation event must be excluded for the evolution of arctic species (Fig. 5.1). The tree and the network were further inspected to characterize the lineages that gave rise to arctic taxa. This was achieved by tracing character states for geographical origin on internal branches on the tree (Fig. 5.1), and by assuming that older sequences tend to occupy central positions in the network (Fig. 5.2). The two sequence analysis methods agree that arctic species originate through several lineages (labeled with a snowflake symbol), up to six according to the phylogenetic tree (Fig. 5.1) and up to eight according to the sequence network (Fig. 5.2). Since some of the informal groups do not have a highly supported internal structure (Fig. 5.1), the number of arctic colonizations might, however, be lower than that predicted (see discussion).

In the tree (Fig. 5.1), the arctic members of section *Orobia* evolved through three different lineages (four in the network, Fig. 5.2), and arctic members of the *Arctobia* section, through two lineages (three in the network). Members (*O. deflexa*) of *Mesogaea* section form the other arctic lineage. The two methods also agree that these arctic lineages evolved from temperate ancestors. Due to lack of resolution in the phylogenetic tree (Fig. 5.1) and incomplete species sampling, the temperate ancestors to arctic species cannot be determined. In the sequence network (Fig. 5.2), nodes comprising arctic species are towards the tip, and according to theory (Huson and Bryant 2006), the more central node to which they connect can be postulated as their ancestor. This logic would imply that, rather than temperate species of the *Orobia* section, a node composed of Central Asian species

of the *Verticillares* and the *Xerobia* sections would be the ancestors of *O. maydelliana*, *O. arctica* var. *barnebyana,* and *O. arctica* var. *koyokukensis* of the *Orobia* section. Similarly, the *O. arctica* var. *arctica* samples from the North American Arctic would have for ancestors North American and Asian members of the *Orobia*, the *Eumorpha* and the *Verticillares* for ancestors (Fig. 5.2). The *O. arctobia* and *O. nigrescens* samples (*Arctobia* section) from the North American Arctic would have the Utah and Montana species *O. besseyi* var. *ventosa* (Welsh 1991) of the *Orobia* section as ancestors, while the arctic *O. podocarpa* from Nunavut would have the Epsilon group composed of Asian and European *Verticillares* species as an ancestor (Fig. 5.2). The *O. deflexa* subsp. *foliolosa* connects to an inferred central node (Fig. 5.2) and is located towards the root of the phylogenetic tree (Fig. 5.1). It should be noted that this taxon, while still considered a low arctic-alpine species, is only marginally present in the Canadian arctic islands (Aiken et al. 2007).

5.5 Discussion

5.5.1 North American arctic *Oxytropis* species evolved independently, more recently than 4.6 million years ago

The two sequence analysis methods used to infer species evolutionary relationships confirm the pattern suggested by *Oxytropis* traditional taxonomy, in which arctic species did not evolve by a single adaptive radiation in the genus, but rather from different temperate lineages. The *Oxytropis* species distributed in the Arctic are from five taxonomic sections (Elven 2011), *Arctobia*, *Orobia*, *Verticillares* (synonymous with *Bacilli*), *Gloeocephala* and *Mesogaea*. All these groups, but less

so for the *Arctobia* section, are mostly composed of temperate species. Thus, arctic *Oxytropis* were hypothesized to have arisen from a minimum of five different lineages.

The phylogenetic tree and the sequences network built here include nine arctic species (out of the 44 that exist over the entire Arctic) from four of these five sections, all from North America. Results show that arctic species evolved from temperate ancestors, through up to six lineages. This estimate may be reduced to four if the lack of internal support for some of these groups is considered. Explicitly, because the Zeta group has no supported internal branches, it cannot be ruled out that its ancestral state was arctic. Although hypothetical, this might merge the evolution of arctic samples of the Zeta group and of *O. podocarpa* into a single event. Additionally, that *O. nigrescens* and *O. arctobia* evolved independently is suspicious because they are infraspecific taxa of a same large species in some classifications (Yurtsev 1997). Even if reduced to four events, this pattern therefore eliminates the hypothesis of an adaptive radiation of arctic *Oxytropis* from a single ancestor. Although dating the nodes for lineages leading to arctic taxa has not been possible, the phylogenetic tree indicates that the only dated *Oxytropis* node (Wojciechowski 2005), which is at 4.6 Ma, happened before the establishment of arctic lineages (the *O. arctobia*, the *O. maydelliana,* the many *O. arctica*, the *O. campestris gracilis* and the *O. podocarpa* lineages) except *O. deflexa* subsp. *foliolosa*. This evidence is, however, not decisive regarding whether these arctic lineages evolved before or after the onset of the arctic climate (3 Ma).

A collection of data from fossil evidence, molecular dating of lineages and species distributions, in combination with the present nuclear sequence data analysis, enable the elaboration of a plausible scenario for *Oxytropis* colonization to the North American Arctic. The fossil record indicates that arctic plants were present in the Arctic 3 Ma (Abbott 2008 and references therein). In this scenario,

Mesogaea-like ancestral taxa evolved first, earlier than 4.6 Ma, probably in Eurasian mountain regions where diversity is high (Malyshev 2008; Langran et al. 2010). Then, many lineages including those that will give rise to arctic taxa diversified, shortly before to shortly after the onset of the arctic climate (3 Ma). Among these taxa, those that possessed the physiological, phenological and morphological characters compatible with life in short and cold growth season were mainly of the *Orobia* and the *Arctobia* sections, and continued to successfully colonize the arctic land cover (Donoghue 2008). Our previous results revealed that some of the molecular characteristics of arctic *Oxytropis* relate to specific gene expression (Archambault and Strömvik 2011) and selective pressures (Archambault and Strömvik 2012b). Limited local speciation could have happened subsequently, such as within the *Arctobia* section (Yurtsev 1997), including in North America. The pattern of an arctic flora mostly formed by parallel evolution of taxa from non-arctic lineages, with only rare and limited evolutionary radiations, was demonstrated to be general in the meta analysis of phylogenies comprising arctic and non-arctic taxa of *Artemisia, Erigeron, Cerastium, Carex, Astragalus, Juncus, Poa, Ranunculus* and many others (Hoffmann and Roser 2009).

5.5.2 *Oxytropis* species are generally cohesive units, with exceptions

The sequence analyses allow for the exploration of additional hypotheses other than the evolution of arctic taxa. Given that *Astragalus* and *Oxytropis* genera are of the same age and have similar geographical centers of origin (Wojciechowski 2005), growth habit, and reproductive system, they could be expected to evolve similar levels of sequence diversity. These two genera are so close to each other that *Oxytropis* was once considered part of *Astragalus*; and they are differentiated by a

single character: the *Oxytropis* beaked keel petals (Welsh 1991). We show, however, that genomes of the various *Oxytropis* species are much more similar to each other than the *Astragalus* species.

Although the sampling was not designed to assess species boundaries, some preliminary observations can be made in this regard. *Oxytropis* individuals generally group with conspecific sequences (with a few exceptions) in both sequence analyses. In most cases where specimens do not strictly group together, the split is neither highly supported in the tree nor are sequences separated by many nodes in the network. The lack of support suggests that these splits may not be evolutionary meaningful. This observed pattern, where molecular data generally follow species boundaries, differs from other studies on rare *Oxytropis* taxa where species distinctions were slightly to considerably blurred. These studies, however, used other types of genetic data such as amplified fragment length polymorphism (AFLP) (Jorgensen et al. 2003; Schonswetter et al. 2004) or random amplified polymorphic DNA (RAPD) (Jorgensen et al. 2003; Artyukova et al. 2004) that can reveal much greater levels of genetic diversity compared to the ITS sequences.

The only cases where ITS sequence analyses indicated highly supported splits are for *O. arctica* and *O. aciphylla* samples. The *O. arctica* situation has previously been analyzed, and it had been concluded that hybridization played a role in refugial areas during the Pleistocene glacial peaks (Jorgensen et al. 2003). The split between *O. aciphylla* samples (Gao et al. 2009) also has a geographic component, but it is still not clear whether blurred genetic boundaries were caused by incomplete lineage sorting or hybridization (Joly et al. 2009). Overall, in *Oxytropis,* the evolutionary processes that can confound relationships did not have general long-term effects.

5.5.3 The *Oxytropis* sequence evolutionary analyses have taxonomic implications, especially for inferring the ancestral group

Taxonomic classifications commonly serve as initial hypotheses for molecular systematics studies. For *Oxytropis*, the circumscription of some problematic groups need to be organized into "aggregates" to reflect uncertainties (Elven 2011), however, the subgeneric and sectional boundaries broadly concord among taxonomic treatments from different continents (Yakovlev et al. 1996; Welsh 2001; Elven 2007; Langran et al. 2010). We further discuss only results from the present study that are supported by more than one representative of a taxon (accessions of one species, or species of a section). This implies that the distant relationship of *Janthina* and *Mesogaea* representatives, although they are of the same subgenus, cannot be critically assessed until an increased sampling is available. In the present study, the relationships suggested by nuclear ribosomal ITS sequences from specimens of almost all the subgenera are in general agreement with traditional taxonomy. The genetic proximity of species of the *Verticillares*, the *Orobia*, the *Gloeocephala* and the *Arctobia* sections shown by the present results parallels their previously noted morphological similarities presented in a phenetic dendrogram (Malyshev 2008). The *O. podocarpa, O. arctobia* and *O. nigrescens* are here only distantly related, although they are all of the *Arctobia* section. A sampling that would comprise an increased number of arctic species of that section, which occur mainly in Alaska, Siberia and the Russian Far East (Elven 2011), could provide preliminary explanations as to why the two Canadian species *O. podocarpa* and *O. arctobia* now have a partially overlapping distribution (Aiken et al. 2007), but are not closely related.

The inclusion of high polyploids, such as the dodecaploid *O. maydelliana*, would be of concern in any molecular phylogenetic analysis. Phylogenetic

reconstructions using the ITS locus are unique, since this locus is formed by hundreds of tandem copies of the transcriptional ribosomal genes unit, where all the copies are assumed to be identical due to gene conversion (Feliner and Rossello 2007). However, gene conversion may not always be completely effective (Alvarez and Wendel 2003). Different populations of an allopolyploid species could theoretically harbour different ITS sequences, especially in cases of recurrent allopolyploidization. The mode of origin of the high polyploid (higher than tetraploid) species included here (*O. maydelliana, O. arctica* var. *arctica, O. arctica* var. *koyokukensis, O. lambertii*) is not yet known. Nevertheless, a reassuring finding is that the different specimens of each of these polyploid taxa remain closely related, including the three *O. maydelliana* that are geographically distributed from Alaska to Baffin Island.

5.5.3.1 The sequence network and the phylogenetic tree disagree for identifying the primitive *Oxytropis* group

Taxonomists proposed that members of the *Phacoxytropis* subgenus are primitive in the genus based on presence of characters also present in primitive *Astragalus* lineages (Yurtsev 1999). This relationship was compatible with early molecular systematic studies in *Astragalus* (Wojciechowski et al. 1993; Wojciechowski et al. 1999), but they included at most eleven *Oxytropis* species. In agreement with this view, the phylogenetic tree we present here clearly identifies members of the *Mesogaea* section as the ancestral lineages of *Oxytropis*. The sequence network topology, however, suggests that members of the *Verticillares* are ancestral within the genus. Considering that this surprising result is at odds with the other evidence, it can be interpreted that the statistical parsimony network can be

misleading for identifying ancestral sequences, especially when applied at the generic level, and when sampling is not yet complete.

Interestingly, an implication of these finding is that the similarities in expression profiles found in our previous study of the two arctic species *O. arctobia* and *O. maydelliana* (Archambault and Strömvik 2011), are indeed not due to a common long evolutionary arctic split from the remainder of the genus, but rather due to common molecular strategies for arctic survival. To conclude, the set of ITS sequences from *Oxytropis* species from different continents and of four subgenera provided sufficient variation to confirm that, as suggested by taxonomic classification, the arctic was colonized by different temperate lineages.

5.6 Acknowledgements

This work was supported by a grant from Natural Sciences and Engineering Research Council of Canada (NSERC) to MVS and by the Centre Sève (Le Fonds québécois de la recherche sur la nature et les technologies, FQRNT). AA was supported by a scholarship from FQRNT, and by Eben-Hopson and Girardin-Vaillancourt fellowships. The authors would like to thank Dr. Susan Aiken, Dr. Laurie Consaul, Dr. Danielle Prévost, Dr. Carolyn Parker, Dr. Tracy Sterling and Plant Gene Resources of Canada (Agriculture and Agri-Food Canada, Saskatoon Research Centre) for providing *Oxytropis* seeds.

Chapter 6

6 General Discussion

Only a very small proportion of the earth's plant species have been able to colonize the Arctic. However, although arctic plants show several characteristics that occur in high frequency in this flora, none are unique. Many of these characteristics are presented in Chapter 2, and include prefloration, rapid growth that resumes rapidly after show melt, extensive root system, absence of seed dormancy and of barriers to germination, common vegetative and asexual reproduction, requirement of longer hours of light to flower, heliotropism, efficient freezing tolerance even during growth season, photosynthesis and respiration rates acclimation to temperature, chloroplast protrusion (increased content in organelles membranes in closer association with mitochondria).

Since gene expression is at the interface of genotype to phenotype, this project was undertaken to characterize the molecular distinguishing features of a set of two arctic plants compared to two relatives from the temperate climate, in terms of gene expression general profile (Hypotheses H2.1 and H2.2), gene regulation under different growth conditions (Hypothesis H3), codons under positive and negative selection (Hypothesis H4) and gene duplication.

The selected model system was the *Oxytropis* genus (Fabaceae) where most species are alpine or lowland temperate, but that also includes many species from the North American Arctic. In order to properly interpret comparative transcriptomics results, genetic divergence and evolutionary relationships within the genus has been estimated (Hypothesis H1.1 and H1.2).

In this chapter, each initial Hypothesis is briefly discussed and then placed in a wider context.

6.1 The *Oxytropis* evolutionary relationships confirm low genetic divergence among species and reveal no adaptive radiation for the arctic species

The commonly used nuclear ribosomal ITS (Internal Transcribed Sequence) nucleotide sequence was gathered for 97 individuals of 30 different *Oxytropis* species that cover a wide range of taxonomical diversity, and sequences were analyzed using a phylogenetic tree and an evolutionary network (Chapter 5) in order to test Hypothesis H1.1 and H1.2. Although the proportion of included species is modest in regard to the size of this large genus (approximately 300 species), the conclusions drawn from the sequence evolutionary analysis are informative.

The average pairwise tree distances for ITS sequences is lower in *Oxytropis* that in its large sister genus, *Astragalus*, which confirms that *Oxytropis* genomes from different species are similar, and that Hypothesis H1.1 (Genetic divergence is low among *Oxytropis* species) cannot be rejected. Sequence evolutionary analyses reveal that different arctic species evolved from many different temperate-like ancestors, and therefore, that Hypothesis H1.2 (Arctic *Oxytropis* do not form a monophyletic group) cannot be rejected.

Such evolutionary pattern was expected given that arctic *Oxytropis* taxa belong to different taxonomic sections (Elven 2007). These two conclusions (Chapter 5) have implications for the *Oxytropis* comparative transcriptomic analysis undertaken here (Chapter 3), as it implies that the library subtraction efficiency was not compromised by using cDNAs from different species, and that specialized molecular features discovered may have evolved relatively independently in the two arctic *Oxytropis* taxa (*O. arctobia* and *O. maydelliana*) under study (Chapter 3).

In combination with other publications that set the timing for the first onset of arctic climate at 3 Ma (Abbott 2008 and references therein), and the timing for one basal node in the *Oxytropis* phylogeny at 4.6 Ma (Wojciechowski et al. 1999; Wojciechowski 2005), a possible scenario for *Oxytropis* arctic species evolution can be envisioned. In this scenario, the earlier *Oxytropis* lineages were *Mesogaea*-like (Chapter 5) from regions of current high species diversity, such as Mongolia and Central Asia (Malyshev 2008). Early lineages later developed into many non arctic-lineages, at around 4.6 Ma (Wojciechowski et al. 1999; Wojciechowski 2005). Shortly after and until 3 Ma, rapid cooling established the arctic growth conditions (Abbott 2008).

Neither the phylogenetic analysis presented here (Chapter 5), nor other evidences can be decisive regarding whether arctic lineages evolved slightly before, or slightly after the onset of the arctic climate. Certainly, it is taxa of the *Orobia* and the *Arctobia* sections that, among ancestral lineages, are those that possessed or developed the physiological, phenological and morphological characters compatible with life in short and cold growth season. Repeated changes in species distribution then occurred, mostly through Beringia (Yurtsev 1999), throughout the last glaciations cycle until 10,000 years ago, during which local speciation happened within the *Arctobia* section (Yurtsev 1999).

A pattern where arctic *Oxytropis* lineages evolved before the establishment of the climate where they are currently distributed is plausible, and would follow the Phylogenetic Niche Conservatism hypothesis (Donoghue 2008 and references therein). This implies that adaptive traits evolved before the advent of a specific climate and is followed by a sorting of species with relevant adaptations at the onset of the novel climatic variations. According to this hypothesis, redistribution is more likely than *in situ* evolution of novel adaptive traits. Changes in distribution in the arctic are facilitated by long distance dispersal, which is common in arctic plants,

according to phylogeographic studies in *Dryas octopetala, Cassiope tetragona* (Alsos et al. 2007) and *Draba subcapitata* (Skrede et al. 2009).

Conclusions from the ITS *Oxytropis* sequence evolutionary analyses have potential taxonomical implications. The different sequence analyses methods disagree in identifying the ancestral *Oxytropis* groups (Chapter 5). The phylogenetic analysis is more trustworthy in the present case, because it is applied at the taxonomic level it was designed for (species-level), it includes an outgroup, and it agrees with current theories on *Oxytropis* evolution (Yurtsev 1999). The phylogenetic analyses identify species of the *Mesogaea* section of the *Phacoxytropis* subgenus as basal lineages within the genus. This agrees with earlier view (Yurtsev 1999), that was based on similarity of *Phacoxytropis* species to the *Phaca* primitive group in the sister genus *Astragalus*. This group was later found to consist in a loose paraphyletic assemblage basal in the genus phylogeny (Wojciechowski et al. 1999; Wojciechowski 2005), which interestingly parallels the tree topology in *Oxytropis* where basal lineages also do not form a monophyletic group.

The analysis from the ITS region sequence data we present (Chapter 5) thus add to the field of *Oxytropis* taxonomy, and to the general understanding of the arctic Flora evolution.

6.2 Transcriptome of plants adapted to abiotic adverse conditions is composed of specialized set of genes

Objectives for comparing transcriptome of arctic and temperate *Oxytropis* species (Chapter 3) were to test that they express different sets of genes from each other in their respective climate conditions (Hypotheses H2.1), and that the biological meaning of this differential gene expression can be understood by

categorizing into gene ontologies (Hypotheses H2.2). The results from library subtractions and Expressed Sequence Tags (ESTs) sequencing between plantlets cDNA from two arctic and two temperate *Oxytropis* species revealed that arctic transcriptomes are enriched in genes of the response to stimulus category, but impoverished in genes of the photosynthesis and of the translation and nucleosome assembly categories. Therefore, Hypotheses H2.1 and H2.2 cannot be rejected.

The field of transcriptome specificity has been examined in several species adapted to stressful conditions, and the *Oxytropis* data we present are unique in that it is, to our knowledge, the first description of a true arctic plant transcriptome. A close example is the *Thellungiella salsuginea* Shandong ecotype from the North Eastern coast of China (wrongly known as *T. halophila,* Amtmann 2009), a cold and salt tolerant species, where the transcriptome was characterized under normal and stress conditions (Inan et al. 2004; Taji et al. 2004; Gong et al. 2005; Taji et al. 2008). A Yukon subarctic ecotype of this species also exists (Griffith et al. 2007), and its transcriptome has been partially characterized under cold, drought and saline conditions (Wong et al. 2005; Wong et al. 2006). In that example, although the *Thellungiella* gene expression profiles were not directly compared to its close relative *Arabidopsis*, quite a few genes expressed had no similarity to the *Arabidopsis* genome, suggesting a certain extent of specialization.

Reviewing the wealth of data existing for comparative gene expression data for the plant species *Thellungiella salsuginea* mainly from the Shandong ecotype, Amtmann (2009) proposed three categories of evolutionary novelties at the transcriptome level for this extremophile: “specificity” (a specific set of genes is regulated in each stress situation); “anticipation” (the transcriptome is constitutively prepared for stress); “lower sensitivity” (higher stresses doses are required to induce transcriptional response).

In most published studies comparing transcriptome of different taxa, including *Oxytropis*, these three novelties are indeed observed. In *Oxytropis*, a specific set of gene categories were detected in the arctic extremophile species, different than in the temperate species transcriptomes when growing in their respective natural climate conditions (Chapter 3). Here and in other studies, the proportion of unknown genes is considerable, for both the taxa adapted to and the one sensitive to adverse conditions (Hammond et al. 2006; Lai et al. 2006; Holliday et al. 2008; Lai et al. 2008; Morinaga et al. 2008; Voelckel et al. 2008; Hegarty et al. 2009). This collection of novel genes certainly contributes to these transcriptome "specificities", but its implications cannot currently be fully assessed. Other transcriptome differences can be categorized as "anticipation" novelties, for instance in the *Malus domestica* resistant to apple-scab (Degenhardt et al. 2005), the Alaskan *Picea* (Holliday et al. 2008) and the zinc-accumulator *Thlaspi* (van de Mortel et al. 2006), where numerous genes of the "stress response" or "defense-related" categories are expressed constitutively, whereas they are upregulated following stress application in other lineages. Expression of the KS-dehydrin gene analyzed in the arctic and temperate *Oxytropis* species, under the two growth conditions, falls within the "anticipation" novelty (Chapter 3). The lower "sensitivity" of *Thellungiella* transcriptome is observed in other species, such as the wheat resistant to fusarium head-blight (Bernardo et al. 2007), but it may not be as widely common, since the opposite trend is seen in ozone resistant *Medicago truncatula* (Puckette et al. 2008), where response to ozone is delayed in the sensitive accession.

It must be recognized that the "anticipation" and "sensitivity" categories of transcriptome novelties are best suited when gene expression is characterized, not only between taxa, but also at different time points, before and after application of the stress under study. In many cases, and in *Oxytropis*, this knowledge is not yet

available, and a "simplicity" category of novelty would best define the transcriptome feature observed in these situations. In many comparative transcriptomics studies among taxa, the taxa with higher resistance to stress or better adaptation to adverse conditions showed a simpler transcriptome (Holliday et al. 2008; Puckette et al. 2009). In the arctic *Oxytropis* species plantlets, not only are there fewer unique genes, they are also from of a narrower set of biological functions. The "simplicity" novelty is not universal, since the trend is not as clear in some species or ecotypes (Hammond et al. 2006; Plessl et al. 2010). The consequence to a "simple" transcriptome as a specialized feature of species adapted to stressful habitat is that downregulated genes may be an important feature of adaptation.

Oxytropis gene expression data (Chapter 3) confirm previous knowledge, but also add novel findings to the general field of comparative transcriptomics. The contrasting difference in gene expression observed for the "response to stimulus" category in *Oxytropis* species parallel change in gene expression previously reported in other extremophiles (Taji et al. 2004; Brosche et al. 2005; Hammond et al. 2006; Knight et al. 2006; van de Mortel et al. 2006; Holliday et al. 2008). For *Oxytropis*, other important differences in expression were found for ribosomal genes. Although changes in gene expression for these genes were not reported as a main feature in other studies, a careful inspection of the differentially expressed genes reveals that in many cases, several ribosomal genes are indeed differentially expressed between species (Filatov et al. 2006; Hammond et al. 2006; Holliday et al. 2008; Voelckel et al. 2008). Given that ribosomal genes may not likely exert a direct control on plant cell and organs arctic resistance, it can be hypothesized that they are part of an increased activation of control mechanisms involved in adaptation to arctic conditions.

6.2.1 Biological meaning of differentially expressed genes; example of the photosynthesis related genes

Obviously, a set of genes with higher or lower expression cannot, by itself, be fully responsible for the complex and subtle adaptations to the Arctic climate. Differences in photosynthesis related gene expression in *Oxytropis* (Chapter 3) or in the Chinese cold tolerant *Thellungiella* (Amtmann 2009) could be part of syndrome that was previously described on differences between leaves and chloroplast anatomy of arctic and temperate plants, which could in turn likely participate in physiological specificities (reviewed in Lutz 2010). In *Oxytropis* from high elevation (Miroslavov and Kravkina 1991) and in *Astragalus* from the Arctic (Miroslavov and Bubolo 1980), leaf cells have more mitochondria and are less vacuolated than the population or species from lower altitudes or latitude. In addition, chloroplasts have smaller grana, with fewer thylakoids, in a pattern that resemble the chloroplast protrusions described for numerous arctic and high alpine plant species (Holzinger et al. 2007; Lutz 2010). Arctic and high alpine plants also typically have thicker palisade cell layers (Miroslavov and Bubolo 1980; Holzinger et al. 2007; Amtmann 2009). These peculiarities of arctic and alpine plants are thought to be involved in physiological differences (Lutz 2010). One example of physiological differences is the respiration rates (Semikhatova et al. 2007) and the photosynthetic rates (Pyankov 1991) for plants from moderate and high latitudes. These rates are similar at ambient temperatures of the plants native habitat. The molecular basis for these anatomical and physiological differences may be encrypted in the list of genes related to photosynthesis found differentially expressed, but their roles could only be understood through a deeper knowledge of their expression profile, precise location and interactions.

6.2.2 The challenges of interpreting a set of differentially expressed genes: the case of stimulus response genes

Moving from a list of differential expressed genes to an understanding of the adaptive value of this difference is still an enormous challenge, as recognized by many in depth reviews of the subject (Storz 2005; Whitehead and Crawford 2006b; Karrenberg and Widmer 2008; Alonso-Blanco et al. 2009; Hodgins-Davis and Townsend 2009; Leakey et al. 2009; Siomos 2009). In organisms with well-characterized physiological genetics, genomes, and population diversity, these metrics could be used to identify a short list of genes with a likely adaptive value among all genes with expression levels polymorphisms (Whitehead and Crawford 2006a; Swindell et al. 2007). Such data are not yet available for any arctic plant species, including *Oxytropis*.

To provide a tentative meaning of biological processes overrepresented in arctic *Oxytropis* plantlets transcriptome, we therefore rely mainly on bibliographic data from agronomical or model plants. Genes related to stress and stimulus response are overrepresented among differentially expressed genes between arctic and temperate *Oxytropis* plantlets. As mentioned, the plasticity in terms of expression variation for genes of this category was previously noted in other plant taxa (Taji et al. 2004; Chen et al. 2005; Degenhardt et al. 2005; Yang and Loopstra 2005; Ageorges et al. 2006; Kliebenstein et al. 2006; van de Mortel et al. 2006; Wang et al. 2006; van Leeuwen et al. 2007; Lai et al. 2008; Puckette et al. 2009; Plessl et al. 2010) and among other organisms, from yeast to fruit flies (Tirosh et al. 2006).

In *Saccharomyces sp.*, to provide tentative biological explanations for the plasticity in expression variation observed in “response to stimulus” genes, a genomic scale association was calculated. It was concluded that genes that vary in

expression tend to be related to “stress and stimulus response” and are more likely to have TATA box containing promoters (Tirosh et al. 2006). The overrepresentation of TATA boxes in promoters of stress response genes could be explained by amplified fluctuations and noises in gene expression brought by particular molecular properties of the TATA boxes, which would allow for expression divergence under neutral drift (Tirosh et al. 2006). The type of promoters for the genes differentially expressed in *Oxytropis* remain to be characterized.

6.3 The Y-segment in KS-dehydrin experienced negative selection

Methods for detecting selection at the sequence level were applied on genomic DNA sequences of four *Oxytropis* species, to test Hypothesis H4 (Chapter 4), that selective pressures are not uniform between genes and along gene regions. This information is highly valuable since the studied genes are still poorly known in *Oxytropis*. Negative selection indicates a protein region under conformational or functional constraints, while sites under positive selection suggest that mutations in the genes sequence were once advantageous, and are therefore promising for exploring novelties and adaptations (Nielsen 2005).

Results from codon selection analyses (Chapter 4) show that selection was not uniform, neither among the genes surveyed, nor along a gene sequence, and therefore, that Hypothesis 4 cannot be rejected. It is however difficult to draw general trends from the pattern of non-uniformity. The only trend suggested by the data is that genes overexpressed in arctic plantlets (PR-10 and KS-dehydrins) are evolutively more constrained at the sequence level (i.e. have overall more codons under negative selection) in arctic species than in other ones.

The non-uniformity is particularly revealing for the KS-dehydrin genes. Although there are codons under negative selection in all surveyed genes (three gene families PR-10, ripening related proteins and KS-dehydrins; and the low copy genes *lhacIII* and *lhcbI*), the proportion is markedly low for the KS-dehydrins. That these proteins evolve relatively freely is in agreement with their unstructured shape (Battaglia et al. 2008) and their proposed function for general binding to lipid membranes (Rahman et al. 2010), partially denatured proteins (Close 1996), or water molecules (Rinne et al. 1999). The fast evolving sequence of KS-dehydrin is a curiosity, but it unfortunately limits the extrapolations that can be made from insightful studies in other species (Rinne et al. 1999; Rorat 2006; Qian et al. 2008; Yang et al. 2009; Rahman et al. 2010). In the Fabaceae, proteins similar to *Oxytropis* dehydrins were described from *Medicago sativa*, *M. truncatula* and *Glycine max* where expression levels increase following many stresses (Takahashi and Shimosaka 1997), especially dehydration (Boudet et al. 2006; Chen et al. 2008a), and where some variants differing in number of repeats provide superior freezing tolerance (Remus-Borel et al. 2010). Non-uniformity in the pattern of codon selection is also evident along the gene sequence. Results showed that within the relatively freely evolving dehydrin sequence, the negatively selected codons cluster within the Y-segment. This is an important novel knowledge, since a biological role has yet to be determined for this region.

6.4 Positive selection is detected mainly in PR-10, in species from contrasting habitats

Results presented here reveal that the pattern of codon under positive selection is not uniform, as it is considerable only in the PR-10, and only in the

complete set of *Oxytropis* sequences, but not in analyses partitioned by arctic and temperate species. That different sequence variants are favoured in these contrasting environments in the PR-10 gene family suggest a shift in the protein use in the different environments. The few codons under positive selection detected in PR-10 of other Fabaceae genera (*Lupinus*, *Pisum*) are in the same protein region as in *Oxytropis*, but rarely at the exact same codon. This is expected given that the hormone binding in PR-10 is flexible, and involves different proteins residues (Mogensen et al. 2002; Koistinen et al. 2005; Pasternak et al. 2005; Fernandes et al. 2008; Fernandes et al. 2009). The PR-10 also differs from other genes surveys, since PR-10 has a higher proportion of negatively selected codons. These high constraints are consistent with suspected dual role of PR-10 (Liu and Ekramoddoullah 2006) in pathogen response by ribonucleolytic activity (Bantignies et al. 2000; Yan et al. 2008) and in plant development by hormone binding (Markovic-Housley et al. 2003; Fernandes et al. 2008). The PR-10 case, where a considerable proportion of codons under both positive and negative selection are detected, compared to other surveyed genes illustrates the usefulness of selection analyses by codon rather than by gene. This feature may not be revealed by an overall d_N:d_S ratio, as in the *Arabidopsis* genome wide analyses of selection in duplicated genes (Warren et al. 2010).

Although the codons under positive selection identified here were not tested functionally, they were for other genes (Barkman et al. 2007; Cavatorta et al. 2008), where the identified sites were proved to carry a genuine phenotypic effect. There are additional instances where the positively selected codons occur in protein regions and sites known for pathogen interaction (Zamora et al. 2009), for prey specialization (Gibbs and Rossiter 2008), or drug resistance (Kosakovsky Pond and Frost 2005b). It is therefore likely that some of the positively selected codons in *Oxytropis* PR-10 will have a biological role; and can serve as a starting point to

better characterize the evolutive implications of these sequence variants for arctic plant biology.

6.5 Response to stimulus genes evolve into gene families of moderate size

The arrangement of the response to stimulus genes into gene families is a finding not covered in the initial hypotheses, but that was further examined in Chapter 4. Phylogenetic tree topologies for PR-10, ripening-related proteins and KS-dehydrins sequences from *Oxytropis* isolated here and other Fabaceae available from public sequences databases revealed that these three gene families are composed of at least two to three copies in the low polyploids and slightly more in the dodecaploid. Furthermore, the number of gene copies of PR-10 appears slightly higher than three when estimated by qPCR copy number experiment, and may show polymorphism in the *Oxytropis* species surveyed here. Our results from four wild *Oxytropis* species did not reveal striking differences in gene copy number for the KS-dehydrins (Chapter 4), even if they exhibit a large difference in expression level among species (Chapter 3) and although distantly related plants carry dehydrins in very different numbers in their genome (e.g. dehydrins in *Hordeum* and *Arabidopsis;* Rorat 2006). In line with the discovery of gene families in *Oxytropis*, the subsequent publication of the genome of the Brassicaceae extremophiles *Thellungiella parvula* (Dassanayake et al. 2011) and of *T. salsuginea* (Wu et al. 2012), revealed that tandem duplications of the stimulus response genes are numerous, and are commonly unique to each species. Indeed, duplication events appear to be a striking feature of all plant genomes (reviewed in Oh et al. 2012).

In *Oxytropis*, data from gene location and genome colinearity is not yet available. In related legume species, however, the PR-10, ripening-related proteins and KS-dehydrin genes were found in one or a few blocks of tandem duplications. The timing of these duplications is not completely resolved here, but given that all sequences of an arctic species did not group into a clade, duplications must predate arctic speciation events. The position of an *Astragalus* (the *Oxytropis* sister genus) ripening-related protein sequence within the *Oxytropis* clade (Chapter 4) suggests that duplications could have occurred prior to the *Oxytropis-Astragalus* split in this gene family.

In *Oxytropis*, the collective results from differential gene expression, codon selection, and gene evolutionary analyses enable the drafting of a series of events that happened in genomes of *Oxytropis* lineages, and lead to arctic species. Duplications occurred for response to stimulus genes in lineages prior to formation of the *Oxytropis* genus, or shortly after for some gene families, but still prior to establishment of major *Oxytropis* lineages. In *Oxytropis*, gene duplications seem not to directly lead to an increase in gene expression, since species that overexpress a gene do not necessarily have additional copies of it. If not brought by gene duplications, the mechanism leading to change in gene expression levels must be of other type, influenced by promoter type for instance. Nevertheless, duplicated gene copies generated material that can be shaped by positive and negative selection, possibly concomitantly to changes in gene expression levels. For some genes, such as PR-10, positive selection might have occurred early on, at the establishment of ancestral arctic and temperate lineages. For other genes, such as KS-dehydrin, selection regimes differ in the arctic and in the temperate species.

Genome evolution through gene duplication, selection at the nucleotide level, and changes in expression levels all contribute to species specificity, which likely allows plants to achieve a phenotype in line with environmental conditions.

Chapter 7

7 Summary

This thesis addresses the hypotheses that among *Oxytropis* species, genetic divergence is low (H1.1) and coding sequences of homologous genes are therefore expected to be conserved among species; and that (H1.2) arctic *Oxytropis* species do not form a monophyletic group within the genus. The nuclear ribosomal internal transcribed spacer (ITS) was isolated and directly sequenced from genomic DNA of 16 *Oxytropis* individuals, mainly of four species (*O. arctobia*, *O. maydelliana*, *O. campestris* subsp. *johannensis*, *O. splendens*). Sequences available from public online databases for 81 additional *Oxytropis* samples were retrieved. The resulting sequence alignment included 30 *Oxytropis* species from a wide taxonomic range, and showed lower average pairwise tree distances compared to that in the large sister genus *Astragalus,* which confirms low genetic divergence in *Oxytropis* (H1.1). Sequences from different individuals of the same species were identical or near identical (with a few exceptions). The *Oxytropis* sequences were analysed using phylogenetic and network approach, and results show that the nine arctic species included evolved probably from six different temperate ancestral lineages. Precisely, the two arctic *Oxytropis* species (*O. arctobia, O. maydelliana*) analysed for transcriptome comparison (Chapter 3) are not related, each rather grouped according to its own taxonomic affinity (Objective 1, H1.2). This indicates that differences in transcriptome composition between arctic and temperate *Oxytropis* plantlets (Chapter 3) are not solely due to a long common evolutionary history.

The hypothesis that arctic and temperate species express different set of genes from each other in their respective natural climate conditions (H2.1) was then tested (Chapter 3). The results from plantlets cDNA library subtraction followed by EST sequencing and annotation of more than 1700 clones resulted in clustering of

ESTs into 121 arctic and 368 temperate genes. Only a very small proportion of the ESTs from each subtracted library showed similarity to an EST in the other subtracted library, demonstrating that the transcriptomes are not composed of the same genes (H2.1). To address the hypothesis that biological and potential adaptive meaning of this differential gene expression can be estimated (H2.2), putative identity of expressed sequences was assessed by similarity search to public online sequences databases, and genes were classified into general categories. These steps showed that genes from the arctic-enriched library are predominantly involved in response to stimulus, whereas genes from the temperate-enriched library are involved in photosynthesis and nucleosome assembly (Objective 2, H2.2). Moreover, both subtracted libraries contained genes involved in ribosome biogenesis and assembly, of different types.

The experimental design undertaken here, where cDNAs from arctic plantlets grown under arctic simulated conditions are compared to cDNAs from temperate plantlets grown under temperate simulated conditions, do not enable the distinction between potentially adaptive features and response to growth conditions. To fill in this gap and test the hypothesis that arctic and temperate *Oxytropis* species regulate these sets of genes differently from each other in different climate conditions (H3); relative expression of four genes, found differentially expressed after Objective 2, were examined by real-time RT-PCR, in plantlets of the four *Oxytropis* species surveyed, under the two (arctic and temperate) climatic conditions (Chapter 3). Expression data for these genes revealed very different responses. The higher expression in the arctic species of one gene, a KS-dehydrin of the "response to stimulus" category, is apparently a specialized feature of arctic plants (H3), since expression was responsive to growth conditions only in the two temperate species, and was similarly elevated under both conditions in the arctic plantlets. This pattern was previously observed in other plants species adapted to stressful conditions. Two

duplicates of the pathogenesis related class 10 (PR-10) gene family may have been recruited differently in the different species: the arctic *O. arctobia* expressed preferentially one gene copy (the PR-10 arctic.contig61), and the three other species expressed the other copy (PR-10 arctic.contig13/36).

The last hypothesis (H4) addressed whether nucleotide sequences have experienced non-uniform levels of negative and positive selective pressures among differentially expressed genes and along gene regions. To test this hypothesis, a total of 96 sequences were isolated from genomic DNA of the four *Oxytropis* species surveyed, for genes among the most differentially expressed (Chapter 3) between arctic and temperate *Oxytropis* plantlets (PR-10, ripening-related proteins and KS-dehydrins gene families and light harvesting photosynthetic apparatus low-copy genes) and codon selection was investigated (Objective 4, H4). Results show that the pattern of nucleotide selection is not uniform neither among genes, nor along the genes sequence, and therefore that the tested hypothesis H4 cannot be rejected. The proportion of negatively selected codon vary among the genes, it is higher in the PR-10, medium in the ripening-related proteins and markedly low in the KS-dehydrins, indicating these latter proteins evolve relatively freely. Furthermore, the few codons under negative selection in the *Oxytropis* KS-dehydrins cluster in the Y-segment, a protein region that has a yet unknown function. Codons under positive selection are not distributed uniformly. They are rare, except for the PR-10 family, where they are noticeably frequent in the complete set of *Oxytropis* PR-10 sequences, but not when PR-10 sequences from arctic and temperate species are analyzed separately. The sequence analysis of *Oxytropis* KS-dehydrins also revealed that they have a K-like – Y_4 – K – S structure, not previously reported.

A feature discovered over the course of genes sequence analyses that was not covered in the initial hypotheses, is their arrangement into gene families. For the three gene families surveyed (PR-10, ripening-related proteins, KS-dehydrins), the

phylogenetic tree topology served as guide to estimate the number of copies (Chapter 4). Results showed that gene families remained modest in size in *Oxytropis* with up to three different genes, including in the hexaploid *O. campestris* subsp. *johannensis*. Only for the high polyploid *O. maydelliana*, there was evidence for a slightly higher number of paralogs. The estimations of copy number from a qPCR experiment also suggest a slightly higher number of copies in the PR-10 family. Tree topology, where all *Oxytropis* sequences formed a clade separate from paralogs of other legume genera, is suggestive of some extent of lineage specific tandem duplication and extinction.

7.1 Future directions

Given the exploratory nature of this project, openings for future research are rich, diverse and numerous. First of all, most of the isolated clones and sequences following the library subtraction are still considered "candidates" for differential expression. The real-time RT-PCR experiment carried out on four genes validated the library subtraction technique, but also revealed that the pattern responsible for the differential expression varies depending on the gene. Unfortunately, at the time being, cost and time associated with the real-time RT-PCR experiments prevent its large-scale use, although it is one of the most precise and trustworthy methods to characterize gene expression. Other confirmation methods, such as hybridization arrays could be envisioned, but are not always as reliable. The massively parallel sequencing methods, commonly called "next generation sequencing", recently gained in popularity and, by providing a high number of short reads, they allow the sequencing of a near complete transcriptome at a reasonable cost. Although these technologies enable transcriptome sequencing at an affordable cost in species without a sequenced genome (Braeutigam and Gowik 2010), the high expected allelic variations in out-crossing species such as *Oxytropis* (Kudo and Harder 2005) combined to still short reads may complicate the assembly step. Furthermore, quantifying the difference in expression may not be accurate in the absence of the sequenced genome, especially for recently duplicated genes, and for genes evolving rapidly at the nucleotide sequence level, as described (Braeutigam and Gowik 2010). This would be of concern for *Oxytropis* species, since some of the differentially expressed genes between species are members of gene families (Archambault and Strömvik 2011) have these features (Archambault and Strömvik 2012b).

There are now many online resources offering access to a wide diversity of data, for instance meta-analyses of gene expression data, and genomic sequences data (including promoters and regulatory elements) across taxa, that could be explored to provide a preliminary understanding of genes differential expression between arctic and temperate *Oxytropis*. For instance, gene expression atlas developed for *Arabidopsis thaliana* (Schmid et al. 2005; Winter et al. 2007) and meta-analyses of microarray (based on Affymetrix GeneChip) expression data (Zimmermann et al. 2004; Wise et al. 2006) could be browsed to distinguish if there is a common profile, in *Arabidopis*, for genes preferentially expressed in arctic or in temperate *Oxytropis* plantlets. This type of search could reveal that arctic-expressed genes tend to be organ-specific or that the temperate-expressed genes are particularly responsive to cold treatments in *Arabidopsis* for example. In addition, exploration of genome information is now facilitated by the PLAZA online platform for plant comparative genomics (Proost et al. 2009) that includes interactive tools to access precompiled data sets covering homologous gene families, multiple sequence alignments, phylogenetic trees, intraspecies whole-genome dot plots, and genomic colinearity between species. Other examples are the databases for plant promoter (Chang et al. 2008; Yamamoto and Obokata 2008), where unfortunately no *Oxytropis* promoter is included so far, that could still be explored to detect common motifs in others species genes homologs for the *Oxytropis* differentially expressed genes.

A research area that needs to be expanded is the characterization of unknown and novel genes. In *Oxytropis*, a quarter of the genes preferentially expressed in the arctic plantlets have no similarity to any other sequences in public sequence databases. Although these genes may be responsible for the long-term colonization of arctic *Oxytropis* species, their exact role has yet to be determined.

It was unexpected to find that genes participating in the ribosome biogenesis and assembly were differentially expressed, some preferentially in the arctic plantlets, some others in the temperate plantlets. Although ribosome roles were initially presumed to be limited to translation of genes into proteins, their central functions in stress and developmental processes now receive more attention (McIntosh and Bonham-Smith 2006), and their role in adaptation to rigorous environmental condition should be explored further in *Oxytropis*.

A similar research area would focus on the "response to stimulus" genes. Indeed, in most of the plant transcriptomics studies comparing gene expression among taxa or lineages, categories related to defense and stress are recognized as differentially expressed, even when the two taxa compared do not primarily differ in their stress response. The "response to stimulus" (Kliebenstein et al. 2006) category includes sub-categories such as "defense" (Hammond et al. 2006; Wang et al. 2006; Hegarty et al. 2009), "cell rescue, defense and virulence" (Chen et al. 2005), or "stress-response" (Kliebenstein et al. 2006; van de Mortel et al. 2006; Holliday et al. 2008; Lai et al. 2008; Voelckel et al. 2008). In most but not all cases, the stress related genes were more highly expressed in the organism adapted to the more stressful conditions. A simplistic shortcut would conclude that defense and stress responses are generally involved in plant adaptation to various abiotic or climatic conditions. However, the explanation might be more subtle. Indeed, as illustrated by the *Oxytropis* PR-10 and ripening related proteins, several of the genes classified in defense or stress categories, may in fact achieve other functions as well. It raises the question whether "stress" and "response to stimulus" Gene Ontology (GO) terms are fully adequate for plant biological processes, or if these terms were applied to genes that once showed an altered (generally an increased) pattern of gene expression following application of a stress. A re-evaluation of this category of genes

seems therefore needed, and may modify the conclusions for many plant comparative transcriptomics studies.

Numerous differentially expressed genes in *Oxytropis* plantlets were members of gene families, this finding warrants a more thorough investigation since lineage-specific expansion and contraction of gene families is thought to play a role in adaptation (Demuth and Hahn 2009). It would be informative to gain additional sequence data for these gene families from more diploids or low-level polyploids *Oxytropis* species. Preliminary data on *Oxytropis* suggested that gene families have not contracted or expanded considerably within the arctic or the temperate lineages (Hahn et al. 2005), but this should be analysed more formally. Expression profiles could then be evaluated for each duplicated copies, in order to detect neofunctionalization or subfunctionalization.

The next steps, concerning the plausible series of genomics events that lead to formation of arctic *Oxytropis* species, will be to determine the causality effect or interrelationships of each of these, and whether they actually provide a selective advantage to arctic species. As mentioned earlier, this will represent a monumental challenge.

Bibliography

Abbott R, Brochmann C (2003) History and evolution of the arctic flora: in the footsteps of Eric Hulten. Molecular Ecology 12(2): 299-313.

Abbott RJ (2008) History, evolution and future of arctic and alpine flora: overview. Plant Ecology & Diversity 1(2): 129-133. 10.1080/17550870802460976.

Adams K (2007) Evolution of duplicate gene expression in polyploid and hybrid plants. Journal of Heredity 98(2): 136-141. 10.1093/jhered/esl061.

Adams KL, Wendel JF (2005) Polyploidy and genome evolution in plants. Current Opinion in Plant Biology 8(2): 135-141. 10.1016/j.pbi.2005.01.001.

Affymetrix (2011) Microarrays - 3' IVT Expression Analysis.

Ageorges A, Fernandez L, Vialet S, Merdinoglu D, Terrier N, Romieu C (2006) Four specific isogenes of the anthocyanin metabolic pathway are systematically co-expressed with the red colour of grape berries. Plant Science 170(2): 372-383.

Aguileta G, Lengelle J, Marthey S, Chiapello H, Rodolphe F, Gendrault A, Yockteng R, Vercken E, Devier B, Fontaine MC, Wincker P, Dossat C, Cruaud C, Couloux A, Giraud T (2010) Finding candidate genes under positive selection in Non-model species: examples of genes involved in host specialization in pathogens. Molecular Ecology 19(2): 292-306. 10.1111/j.1365-294X.2009.04454.x.

Aiken S, Nature CMo, Canada NRC (2007) Flora of the Canadian Arctic Archipelago. In. NRC Research Press, Ottawa, pp 1 CD-ROM

Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M (2009) What has natural variation taught us about plant development, physiology, and adaptation? Plant Cell 21(7): 1877-1896. 10.1105/tpc.109.068114.

Alsos IG, Eidesen PB, Ehrich D, Skrede I, Westergaard K, Jacobsen GH, Landvik JY, Taberlet P, Brochmann C (2007) Frequent long-distance plant colonization in the changing Arctic. Science 316(5831): 1606-1609. 10.1126/science.1139178.

Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25(17): 3389-3402.

Alvarez I, Wendel J (2003) Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetics and Evolution 29(3): 417-434.

Amtmann A (2009) Learning from evolution: *Thellungiella* generates new knowledge on essential and critical components of abiotic stress tolerance in plants. Molecular Plant 2(1): 3-12. 10.1093/mp/ssn094.

Amunts A, Toporik H, Borovikova A, Nelson N (2010) Structure determination and improved model of plant photosystem I. Journal of Biological Chemistry 285(5): 3478-3486. 10.1074/jbc.M109.072645.

Anisimova M, Liberles DA (2007) The quest for natural selection in the age of comparative genomics. Heredity 99(6): 567-579. 10.1038/sj.hdy.6801052.

Archambault A, Strömvik MV (2011) PR-10, defensin and cold dehydrin genes are among those over expressed in *Oxytropis* (Fabaceae) species adapted to the Arctic. Functional & Integrative Genomics 11(3): 497-505. 10.1007/s10142-011-0223-6.

Archambault A, Strömvik MV (2012a) Evolutionary relationships in *Oxytropis* species, as estimated from the nuclear ribosomal internal transcribed spacer (ITS) sequences point to multiple expansions into the Arctic. Botany-Botanique 90(8): 770-779. 10.1139/b2012-023.

Archambault A, Strömvik MV (2012b) The Y-segment of novel cold dehydrin genes is conserved and codons in the PR-10 genes are under positive selection in *Oxytropis* (Fabaceae) from contrasting climates. Molecular Genetics and Genomics 287(2): 123-142. 10.1007/s00438-011-0664-6.

Arft A, Walker M, Gurevitch J, Alatalo J, Bret-Harte M, Dale M, Diemer M, Gugerli F, Henry G, Jones M, Hollister R, Jonsdottir I, Laine K, Levesque E, Marion G, Molau U, Molgaard P, Nordenhall U, Raszhivin V, Robinson C, Starr G, Stenstrom A, Stenstrom M, Totland O, Turner P, Walker L, Webber P, Welker J, Wookey P (1999) Responses of tundra plants to experimental warming: Meta-analysis of the international tundra experiment. Ecological Monographs 69(4): 491-511.

Armisen D, Lecharny A, Aubourg S (2008) Unique genes in plants: specificities and conserved features throughout evolution. BMC Evolutionary Biology 8: 280. 10.1186/1471-2148-8-280.

Artyukova E, Kholina A, Kozyrenko M, Zhuravlev Y (2004) Analysis of genetic variation in rare endemic species *Oxytropis chankaensis* Jurtz. (Fabaceae) using RAPD markers. Russian Journal of Genetics 40(7): 710-716.

Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry J, Davis A, Dolinski K, Dwight S, Eppig J, Harris M, Hill D, Issel-Tarver L, Kasarskis A, Lewis S, Matese J, Richardson J, Ringwald M, Rubin G, Sherlock G (2000) Gene Ontology: tool for the unification of biology. Nature Genetics 25(1): 25-29.

Bantignies B, Seguin J, Muzac I, Dedaldechamp F, Gulick P, Ibrahim R (2000) Direct evidence for ribonucleolytic activity of a PR-10-like protein from white lupin roots. Plant Molecular Biology 42(6): 871-881.

Barakat A, Szick-Miranda K, Chang IF, Guyot R, Blanc G, Cooke R, Delseny M, Bailey-Serres J (2001) The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome. Plant Physiology 127(2): 398-415. 10.1104/pp.010265.

Barkman TJ, Martins TR, Sutton E, Stout JT (2007) Positive selection for single amino acid change promotes substrate discrimination of a plant volatile-producing enzyme. Molecular Biology and Evolution 24(6): 1320-1329. 10.1093/molbev/msm053.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. Trends in Ecology & Evolution 23(1): 38-44. 10.1016/j.tree.2007.09.008.

Barrett T, Troup D, Wilhite S, Ledoux P, Rudnev D, Evangelista C, Kim I, Soboleva A, Tomashevsky M, Edgar R (2007) NCBI GEO: mining tens of millions of expression profiles - database and tools update. Nucleic Acids Research 35: D760-D765.

Battaglia M, Olvera-Carrillo Y, Garciarrubio A, Campos F, Covarrubias AA (2008) The enigmatic LEA proteins and other hydrophilins. Plant Physiology 148(1): 6-24. 10.1104/pp.108.120725.

Bell K, Bliss L (1977) Overwinter phenology of plants in a polar semi-desert. Arctic 30: 118-121.

Bell K, Bliss L (1980) Plant reproduction in a High Arctic environment. Arctic and Alpine Research 12(1): 1-10.

Berberich T, Uebeler M, Feierabend J (2000) cDNA cloning of cytoplasmic ribosomal protein S7 of winter rye (*Secale cereale*) and its expression in low-temperature-treated leaves. Biochimica et Biophysica Acta - Gene Structure and Expression 1492(1): 276-279.

Beritognolo I, Sabatti M, Brosche M, Scarascia Mugnozza G (2008) Functional genomics to discover genes for salt tolerance in annual and perennial plants. In: Abdelly C, Ozturk M, Ashraf M, Grignon C (eds) Biosaline Agriculture and High Salinity Tolerance. Birkhauser Verlag, Switzerland, pp 273-286

Berkner H, Neudecker P, Mittag D, Ballmer-Weber B, Schweimer K, Vieths S, Rosch P (2009) Cross-reactivity of pollen and food allergens: soybean Gly m 4 is a member of the Bet v 1 superfamily and closely resembles yellow lupine proteins. Bioscience Reports 29(0): 183-192. DOI 10.1042/BSR20080117.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Research 28(1): 235-242.

Bernardo A, Bai GH, Guo PG, Xiao K, Guenzi AC, Ayoubi P (2007) *Fusarium graminearum*-induced changes in gene expression between Fusarium head blight-resistant and susceptible wheat cultivars. Functional & Integrative Genomics 7(1): 69-77. 10.1007/s10142-006-0028-1.

Bies-Etheve N, Gaubier-Comella P, Debures A, Lasserre E, Jobet E, Raynal M, Cooke R, Delseny M (2008) Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. Plant Molecular Biology 67(1-2): 107-124. 10.1007/s11103-008-9304-x.

Biesiadka J, Bujacz G, Sikorski M, Jaskolski M (2002) Crystal structures of two homologous pathogenesis-related proteins from yellow lupine. Journal of Molecular Biology 319(5): 1223-1234. DOI 10.1016/S0022-2836(02)00385-6.

Billings W (1987) Constraints to plant growth, reproduction, and establishment in arctic environments. Arctic and Alpine Research 19(4): 357-365.

Billings W, Mooney H (1968) Ecology of arctic and alpine plants. Biological Reviews of the Cambridge Philosophical Society 43: 481-529.

Bliss L, Gold W (1999) Vascular plant reproduction, establishment, and growth and the effects of cryptogamic crusts within a polar desert ecosystem, Devon Island, NWT, Canada. Canadian Journal of Botany - Revue Canadienne de Botanique 77(5): 623-636.

Bliss LC (1971) Arctic and alpine plant life cycles. Annual Review of Ecology and Systematics 2(ArticleType: research-article / Full publication date: 1971 / Copyright © 1971 Annual Reviews): 405-438.

Blüthgen N, Karsten B, Branka C, Maciej S, Hanspeter H, Dieter B (2005) Biological profiling of gene groups utilizing Gene Ontology. Genome Informatics 16(1): 106-115.

Boudet J, Buitink J, Hoekstra F, Rogniaux H, Larre C, Satour P, Leprince O (2006) Comparative analysis of the heat stable proteome of radicles of *Medicago truncatula* seeds during germination identifies late embryogenesis abundant proteins associated with desiccation tolerance. Plant Physiology 140(4): 1418-1436.

Braeutigam A, Gowik U (2010) What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. Plant Biology 12(6): 831-841. 10.1111/j.1438-8677.2010.00373.x.

Braun K, Romero J, Liddell C, Creamer R (2003) Production of swainsonine by fungal endophytes of locoweed. Mycological Research 107(8): 980-988. 10.1017/S095375620300813X.

Broadley M, White P, Hammond J, Graham N, Bowen H, Emmerson Z, Fray R, Iannetta P, McNicol J, May S (2008) Evidence of neutral transcriptome evolution in plants. New Phytologist 180(3): 587-593.

Brochmann C, Brysting A, Alsos I, Borgen L, Grundt H, Scheen A, Elven R (2004) Polyploidy in arctic plants. Biological Journal of the Linnean Society 82(4): 521-536.

Brochmann C, Brysting AK (2008) The Arctic – an evolutionary freezer? Plant Ecology & Diversity 1(2): 181 - 195.

Broderick K, Pittock C, Arioli T, Creaser E, Weinman J, Rolfe B (1997) Pathogenesis-related proteins in *Trifolium subterraneum*: A general survey and subsequent characterisation of a protein inducible by ethephon and redlegged earth mite attack. Australian Journal of Plant Physiology 24(6): 819-829.

Brosche M, Vinocur B, Alatalo E, Lamminmaki A, Teichmann T, Ottow E, Djilianov D, Afif D, Bogeat-Triboulot M, Altman A, Polle A, Dreyer E, Rudd S, Lars P, Auvinen P, Kangasjarvi J (2005) Gene expression and metabolite profiling of *Populus euphratica* growing in the Negev desert. Genome Biology 6(12).

Bubner B, Baldwin IT (2004) Use of real-time PCR for determining copy number and zygosity in transgenic plants. Plant Cell Reports 23(5): 263-271. 10.1007/s00299-004-0859-y.

Buchner O, Holzinger A, Lutz C (2007) Effects of temperature and light on the formation of chloroplast protrusions in leaf mesophyll cells of high alpine plants. Plant Cell and Environment 30(11): 1347-1356. 10.1111/j.1365-3040.2007.01707.x.

Callaghan T, Bjorn L, Chernov Y, Chapin T, Christensen T, Huntley B, Ims R, Johansson M, Jolly D, Jonasson S, Matveyeva N, Panikov N, Oechel W, Shaver G, Elster J, Henttonen H, Laine K, Taulavuori K, Taulavuori E, Zockler C (2004) Biodiversity, distributions and adaptations of arctic species in the context of environmental change. Ambio 33(7): 404-417.

Callaghan TV, Bj√∂rn LO, Chapin III FS, Chernov Y, Christensen TR, Huntley B, Ims R, Johansson M, Riedlinger DJ, Jonasson S, Matveyeva N, Oechel W, Panikov N, Shaver G (2005) Arctic tundra and polar desert ecosystems. In: Arctic C, Arctic Climate Impact A, Arctic M, Assessment P, Program for the Conservation of Arctic Flora and F, International Arctic Science C (eds) Arctic Climate Impact Assessment. Cambridge University Press, Cambridge; New York, pp 243-352

Cannon S, Mitra A, Baumgarten A, Young N, May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biology 4(1): 10.

Carr AC, Moore SD (2012) Robust quantification of polymerase chain reactions using global fitting. PLoS ONE 7(5): e37640. 10.1371/journal.pone.0037640.

Carvalho AD, Gomes VM (2009) Plant defensins - Prospects for the biological functions and biotechnological properties. Peptides 30(5): 1007-1020. 10.1016/j.peptides.2009.01.018.

Casati P, Stapleton A, Blum J, Walbot V (2006) Genome-wide analysis of high-altitude maize and gene knockdown stocks implicates chromatin remodeling proteins in response to UV-B. The Plant Journal 46(4): 613-627.

Cavatorta JR, Savage AE, Yeam I, Gray SM, Jahn MM (2008) Positive Darwinian selection at single amino acid sites conferring plant virus resistance. Journal of Molecular Evolution 67(5): 551-559. 10.1007/s00239-008-9172-7.

Cervigni G, Paniego N, Pessino S, Selva J, Diaz M, Spangenberg G, Echenique V (2008) Gene expression in diplosporous and sexual *Eragrostis curvula* genotypes with differing ploidy levels. Plant Molecular Biology 67(1-2): 11-23.

Chague V, Just J, Mestiri I, Balzergue S, Tanguy AM, Huneau C, Huteau V, Belcram H, Coriton O, Jahier J, Chalhoub B (2010) Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. New Phytologist 187(4): 1181-1194. 10.1111/j.1469-8137.2010.03339.x.

Chang WC, Lee TY, Huang HD, Huang HY, Pan RL (2008) PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. BMC Genomics 9. 10.1186/1471-2164-9-561.

Chapin F (1983) Direct and indirect effects of temperature on arctic plants. Polar Biology 2: 47-52.

Chelaifa H, Mahe F, Ainouche M (2010a) Transcriptome divergence between the hexaploid salt-marsh sister species *Spartina maritima* and *Spartina alterniflora* (Poaceae). Molecular Ecology 19(10): 2050-2063. 10.1111/j.1365-294X.2010.04637.x.

Chelaifa H, Monnier A, Ainouche M (2010b) Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina x townsendii* and *Spartina anglica* (Poaceae). New Phytologist 186(1): 161-174. 10.1111/j.1469-8137.2010.03179.x.

Chen D, Liang M, DeWald D, Weimer B, Peel M, Bugbee B, Michaelson J, Davis E, Wu Y (2008a) Identification of dehydration responsive genes from two non-nodulated alfalfa cultivars using *Medicago truncatula* microarrays. Acta Physiologiae Plantarum 30(2): 183-199.

Chen L, Zhang SM, Illa E, Song LJ, Wu SD, Howad W, Arus P, van de Weg E, Chen KS, Gao ZS (2008b) Genomic characterization of putative allergen genes in peach/almond and their synteny with apple. BMC Genomics 9. 10.1186/1471-2164-9-543.

Chen W, Chang S, Hudson M, Kwan W, Li J, Estes B, Knoll D, Shi L, Zhu T (2005) Contribution of transcriptional regulation to natural variations in *Arabidopsis*. Genome Biology 6(4): -.

Cheng C, Yun K, Ressom H, Mohanty B, Bajic V, Jia Y, Yun S, de los Reyes B (2007) An early response regulatory cluster induced by low temperature and hydrogen peroxide in seedlings of chilling-tolerant japonica rice. BMC Genomics 8: 175. doi:10.1186/1471-2164-8-175.

Chung M, Gelembiuk G, Givnish T (2004) Population genetics and phylogeography of endangered *Oxytropis campestris* var. *chartacea* and relatives: arctic-alpine disjuncts in eastern North America. Molecular Ecology 13(12): 3657-3673. DOI 10.1111/j.1365-294X.2004.02360.x.

Clark R, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu T, Fu G, Hinds D, Chen H, Frazer K, Huson D, Schoelkopf B, Nordborg M, Raetsch G, Ecker J, Weigel D (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. Science 317: 338-342. 10.1126/science.1138632.

Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. Molecular Ecology 9(10): 1657-1659.

Close T (1996) Dehydrins: Emergence of a biochemical role of a family of plant dehydration proteins. Physiologia Plantarum 97(4): 795-803.

Collective (2012) Adaptation. In: Wikipedia, http://en.wikipedia.org/wiki/Adaptation

Conesa A, Gotz S, Garcia-Gomez J, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18): 3674-3676.

Cook D, Gardner D, Ralphs M, Pfister J, Welch K, Green B (2009) Swainsoninine concentrations and endophyte amounts of *Undifilum oxytropis* in different plant parts of *Oxytropis sericea*. Journal of Chemical Ecology 35(10): 1272-1278. DOI 10.1007/s10886-009-9710-9.

Dassanayake M, Oh DH, Haas JS, Hernandez A, Hong H, Ali S, Yun DJ, Bressan RA, Zhu JK, Bohnert HJ, Cheeseman JM (2011) The genome of the extremophile crucifer *Thellungiella parvula*. Nature Genetics 43(9): 913-U137. 10.1038/ng.889.

Degenhardt J, Al-Masri A, Kurkcuoglu S, Szankowski I, Gau A (2005) Characterization by suppression subtractive hybridization of transcripts that are differentially expressed in leaves of apple scab-resistant and susceptible cultivars of *Malus domestica*. Molecular Genetics and Genomics 273(4): 326-335.

Demuth JP, Hahn MW (2009) The life and death of gene families. Bioessays 31(1): 29-39. 10.1002/bies.080085.

Diatchenko L, Lau Y, Campbell A, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov E, Siebert P (1996) Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proceedings of the National Academy of Sciences of the United States of Americs 93(12): 6025-6030.

Diatchenko L, Lukyanov S, Lau Y, Siebert P (1999) Suppression subtractive hybridization: A versatile method for identifying differentially expressed genes. CDNA Preparation and Characterization 303: 349-380.

Donoghue MJ (2008) A phylogenetic perspective on the distribution of plant diversity. Proceedings of the National Academy of Sciences of the United States of America 105: 11549-11555. 10.1073/pnas.0801962105.

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. Annual Review of Genetics 42: 443-461. 10.1146/annurev.genet.42.110807.091524.

Drummond AJ, Ashton B, M. C, J. H, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A (2008) Geneious. In, 4.0 edn. Biomatters, Auckland, NZ

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7: 214. 10.1186/1471-2148-7-214.

Duarte J, Wall PK, Edger P, Landherr L, Ma H, Pires JC, Leebens-Mack J, dePamphilis C (2010) Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. BMC Evolutionary Biology 10(1): 61.

Duvick J, Fu A, Muppirala U, Sabharwal M, Wilkerson MD, Lawrence CJ, Lushbough C, Brendel V (2008) PlantGDB: a resource for comparative plant genomics. Nucleic Acids Research 36: D959-D965. 10.1093/nar/gkm1041.

Eckert AJ, Bower AD, Gonzalez-Martinez SC, Wegrzyn JL, Coop G, Neale DB (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). Molecular Ecology 19(17): 3789-3805. 10.1111/j.1365-294X.2010.04698.x.

Edger PP, Pires JC (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Research 17(5): 699-717. 10.1007/s10577-009-9055-9.

Elmer KR, Fan S, Gunter HM, Jones JC, Boekhoff S, Kuraku S, Meyer A (2010) Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. Molecular Ecology 19: 197-211. Doi 10.1111/J.1365-294x.2009.04488.X.

Elven R (2007) Checklist of the Panarctic Flora (PAF) Vascular Plants. In. Russian Foundation for Basic Research Grant N 07-04-01015-a

Elven R (2011) Annotated Checklist of the Panarctic Flora (PAF) Vascular plants. In, Natural History Museum, University of Oslo

Engelmann S, Wiludda C, Burscheidt J, Gowik U, Schlue U, Koczor M, Streubel M, Cossu R, Bauwe H, Westhoff P (2008) The gene for the P-subunit of glycine decarboxylase from the C-4 species *Flaveria trinervia*: Analysis of transcriptional control in transgenic *Flaveria bidentis* (C-4) and *Arabidopsis* (C-3). Plant Physiology 146(4): 1773-1785.

Falgueras J, Lara AJ, Canton FR, Perez-Trabado G, Claros MG (2007) SeqTrim - A validation and trimming tool for all purpose sequence reads. In: Corchado E, Corchado JM, Abraham A (eds) Innovations in Hybrid Intelligent Systems. Springer, Berlin / Heidelberg, pp 353-360

Feliner G, Rossello J (2007) Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. Molecular Phylogenetics and Evolution: 911-919. 10.1016/j.ympev.2007.01.013.

Fernandes H, Bujacz A, Bujacz G, Jelen F, Jasinski M, Kachlicki P, Otlewski J, Sikorski M, Jaskolski M (2009) Cytokinin-induced structural adaptability of a *Lupinus luteus* PR-10 protein. FEBS Journal 276: 1596-1609. DOI 10.1111/j.1742-4658.2009.06892.x.

Fernandes H, Pasternak O, Bujacz G, Bujacz A, Sikorski M, Jaskolski M (2008) *Lupinus luteus* pathogenesis-related protein as a reservoir for cytokinin. Journal of Molecular Biology 378(5): 1040-1051.

Filatov V, Dowdle J, Smirnoff N, Ford-Lloyd B, Newbury H, Macnair M (2006) Comparison of gene expression in segregating families identifies genes and genomic regions involved in a novel adaptation, zinc hyperaccumulation. Molecular Ecology 15(10): 3045-3059.

Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. New Phytologist 183(3): 557-564. 10.1111/j.1469-8137.2009.02923.x.

Fluch S, Olmo CC, Tauber S, Stierschneider M, Kopecky D, Reichenauer TG, Matusikova I (2008) Transcriptomic changes in wind-exposed poplar leaves are dependent on developmental stage. Planta 228(5): 757-764. 10.1007/s00425-008-0777-2.

Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annual Review of Plant Biology 60: 433-453. 10.1146/annurev.arplant.043008.092122.

Gao J, Lu P, Wang J-N, Jin F (2009) Molecular phylogeny of several species of *Oxytropis* DC. Based on 5.8S rDNA / ITS Sequence. Acta Agriculturae Boreali-Sinica 06. CNKI:SUN:HBNB.0.2009-06-036.

Gao ZS, van de Weg WE, Schaart JG, Schouten HJ, Tran DH, Kodde LP, van der Meer IM, van der Geest AHM, Kodde J, Breiteneder H, Hoffmann-Sommergruber K, Bosch D, Gilissen L (2005) Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (*Malus domestica*). Theoretical and Applied Genetics 111(1): 171-183. 10.1007/s00122-005-2018-4.

Gauthier G, Rochefort L, Reed A (1996) The exploitation of wetland ecosystems by herbivores on Bylot island. Geoscience Canada 23(4): 253-259.

Gibbs HL, Rossiter W (2008) Rapid evolution by positive selection and gene gain and loss: PLA_2 venom genes in closely related *Sistrurus* rattlesnakes with divergent diets. Journal of Molecular Evolution 66(2): 151-166. 10.1007/s00239-008-9067-7.

Goldblatt P (2007) The Index to Plant Chromosome Numbers - past and future. Taxon 56(4): 984-986.

Gong QQ, Li PH, Ma SS, Rupassara SI, Bohnert HJ (2005) Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. Plant Journal 44(5): 826-839. 10.1111/j.1365-313X.2005.02587.x.

Goodstein DM, Shu SQ, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research 40(D1): D1178-D1186. 10.1093/nar/gkr944.

Gossmann TI, Song BH, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV, Filatov DA, Eyre-Walker A (2010) Genome wide analyses reveal little evidence for adaptive evolution in many plant species. Molecular Biology and Evolution 27(8): 1822-1832. 10.1093/molbev/msq079.

Gotz S, Garcia-Gomez J, Terol J, Williams T, Nagaraj S, Nueda M, Robles M, Talon M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research 36(10): 3420-3435.

Goulas E, Richard-Molard C, Le Dily F, Le Dantec C, Ozouf J, Ourry A (2007) A cytosolic vegetative storage protein (TrVSP) from white clover is encoded by a cold-inducible gene. Physiologia Plantarum 129(3): 567-577.

Green P (2002) Phred, Phrap, and Consed. In. University of Washington, Seattle

Griffith M, Timonin M, Wong ACE, Gray GR, Akhter SR, Saldanha M, Rogers MA, Weretilnyk EA, Moffatt B (2007) *Thellungiella*: an *Arabidopsis*-related model plant adapted to cold temperatures. Plant Cell and Environment 30(5): 529-538. 10.1111/j.1365-3040.2007.01653.x.

Grulke N, Bliss L (1988) Comparative life-history characteristics of two High Arctic grasses, Northwest Territories. Ecology 69(2): 484-496.

Grundt H, Kjolner S, Borgen L, Rieseberg L, Brochmann C (2006) High biological species diversity in the arctic flora. Proceedings of the National Academy of Sciences of the United States of Americs 103(4): 972-975.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52(5): 696-704. 10.1080/10635150390235520.

Guo HY, Wang WW, Yang N, Guo BL, Zhang S, Yang RJ, Yuan Y, Yu JL, Hu SN, Sun QS, Yu J (2010) DNA barcoding provides distinction between Radix Astragali and its adulterants. Science China-Life Sciences 53(8): 992-999. 10.1007/s11427-010-4044-y.

Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N (2005) Estimating the tempo and mode of gene family evolution from comparative genomic data. Genome Research 15(8): 1153-1160.

Hammond J, Bowen H, White P, Mills V, Pyke K, Baker A, Whiting S, May S, Broadley M (2006) A comparison of the *Thlaspi caerulescens* and *Thlaspi arvense* shoot transcriptomes. New Phytologist 170(2): 239-260.

Handschuh L, Femiak I, Kasperska A, Figlerowicz M, Sikorski M (2007) Structural and functional characteristics of two novel members of pathogensis-related multigene family of class 10 from yellow lupine. Acta Biochimica Polonica 54: 783-796.

Hanks J, Snyder A, Graham M, Shah R, Blaylock L, Harrison M, Shah D (2005) Defensin gene family in *Medicago truncatula*: structure, expression and induction by signal molecules. Plant Molecular Biology 58(3): 385-399.

Hannah M, Heyer A, Hincha D (2005) A global survey of gene regulation during cold acclimation in *Arabidopsis thaliana*. Plos Genetics 1(2): 179-196.

Hegarty M, Barker G, Brennan A, Edwards K, Abbott R, Hiscock S (2009) Extreme changes to gene expression associated with homoploid hybrid speciation. Molecular Ecology 18(5): 877-889.

Hellsten U, Mitros T, Prochnik S, Rokhsar D (2010) Phytozome 5.0. University of California Regents; Joint Genome Institute; Center for Integrative Genomics;. http://www.phytozome.net/Phytozome_info.php. Accessed.

Henry G, Molau U (1997) Tundra plants and climate change: the International Tundra Experiment (ITEX). Global Change Biology 3: 1-9.

Higgins KL, Arnold AE, Miadlikowska J, Sarvate SD, Lutzoni F (2007) Phylogenetic relationships, host affinity, and geographic structure of boreal and arctic endophytes from three major plant lineages. Molecular Phylogenetics and Evolution 42(2): 543-555.

Hino T, Kanamori E, Shen J, Kouyama T (2004) An icosahedral assembly of the light-harvesting chlorophyll a/b protein complex from pea chloroplast thylakoid membranes. Acta Crystallographica Section D-Biological Crystallography D60: 803-809. 10.1107/S0907444904003233.

Hodgins-Davis A, Townsend JP (2009) Evolving gene expression: from G to E to G x E. Trends in Ecology & Evolution 24(12): 649-658. 10.1016/j.tree.2009.06.011.

Hoffmann M, Roser M (2009) Taxon recruitment of the arctic flora: an analysis of phylogenies. New Phytologist 182(3): 774-780.

Holliday J, Ralph S, White R, Bohlmann J, Aitken S (2008) Global monitoring of autumn gene expression within and among phenotypically divergent populations of Sitka spruce (*Picea sitchensis*). New Phytologist 178(1): 103-122.

Holmen K (1962) Chromosome studies in some arctic Alaskan Leguminosae. Botanical Notes 115: 87–92.

Holzinger A, Wasteneys G, Lutz C (2007) Investigating cytoskeletal function in chloroplast protrusion formation in the arctic-alpine plant *Oxyria digyna*. Plant Biology 9(3): 400-410.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17(8): 754-755.

Huh G, Matsuura Y, Meshi T, Iwabuchi M (1995) Differential expression of the two types of histone H2A genes in wheat. Biochimica et Biophysica Acta - Gene Structure and Expression 1261(1): 155-160.

Hulten E (1968) Flora of Alaska and neighboring territories; a manual of the vascular plants. Stanford University Press, Stanford, CAL

Hundertmark M, Hincha DK (2008) LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. BMC Genomics 9. 10.1186/1471-2164-9-118.

Huson D, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23(2): 254-267.

Inan G, Zhang Q, Li P, Wang Z, Cao Z, Zhang H, Zhang C, Quist T, Goodwin S, Zhu J, Shi H, Damsz B, Charbaji T, Gong Q, Ma S, Fredricksen M, Galbraith D, Jenks M, Rhodes D, Hasegawa P, Bohnert H, Joly R, Bressan R, Zhu J (2004) Salt cress. A halophyte and cryophyte *Arabidopsis* relative model system and its applicability to molecular genetic analyses of growth and development of extremophiles. Plant Physiology 135(3): 1718-1737.

Ishibashi N, Yamauchi D, Minamikawa T (1990) Stored messenger-RNA in cotyledons of *Vigna unguilculata* seeds - nucleotide sequence of cloned cDNA for a stored messenger RNA and induction of its synthesis by precocious germination. Plant Molecular Biology 15(1): 59-64.

Iturriaga E, Leech M, Barratt D, Wang T (1994) Two ABA-responsive proteins from pea (*Pisum sativum* L.) are closely related to intracellular pathogenesis-related proteins. Plant Molecular Biology 24(1): 235-240.

Jensen JD, Wong A, Aquadro CF (2007) Approaches for identifying targets of positive selection. Trends in Genetics 23(11): 568-577. 10.1016/j.tig.2007.08.009.

Joly S, McLenachan PA, Lockhart PJ (2009) A statistical approach for distinguishing hybridization and incomplete lineage sorting. American Naturalist 174(2): E54-E70. 10.1086/600082.

Joly S, Stevens MI, van Vuuren BJ (2007) Haplotype networks can be misleading in the presence of missing data. Systematic Biology 56(5): 857-862. 10.1080/10635150701633153.

Jorgensen J, Stehlik I, Brochmann C, Conti E (2003) Implications of ITS sequences and RAPD markers for the taxonomy and biogeography of the *Oxytropis campestris* and *O. arctica* (Fabaceae) complexes in Alaska. American Journal of Botany 90(10): 1470-1480.

Juenger TE, Sen S, Bray E, Stahl E, Wayne T, McKay J, Richards JH (2010) Exploring genetic and expression differences between physiologically extreme ecotypes: comparative genomic hybridization and gene expression studies of Kas-1 and Tsu-1 accessions of *Arabidopsis thaliana*. Plant Cell and Environment 33(8): 1268-1284. 10.1111/j.1365-3040.2010.02146.x.

Jump AS, Hunt JM, Martinez-Izquierdo JA, Penuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. Molecular Ecology 15(11): 3469-3480. 10.1111/j.1365-294X.2006.03027.x.

Junttila O, Robberecht R (1993) The influence of season and phenology on freezing tolerance in *Silene acaulis* L, a sub-arctic and arctic cushion plant of circumpolar distribution. Annals of Botany 71(5): 423-426.

Kane NC, Rieseberg LH (2007) Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. Genetics 175(4): 1823-1834. 10.1534/genetics.106.067728.

Kane NC, Rieseberg LH (2008) Genetics and evolution of weedy *Helianthus annuus* populations: adaptation of an agricultural weed. Molecular Ecology 17(1): 384-394. 10.1111/j.1365-294X.2007.03467.x.

Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 28(1): 27-30.

Kapros T, Bogre L, Nemeth K, Bako L, Gyorgeyey J, Wu S, Dudits D (1992) Differential expression of histone H3 gene variants during cell cycle and somatic embryogenesis in Alfalfa. Plant Physiology 98(2): 621-625.

Karrenberg S, Widmer A (2008) Ecologically relevant genetic variation from a non-*Arabidopsis* perspective. Current Opinion in Plant Biology 11(2): 156-162. 10.1016/j.pbi.2008.01.004.

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30(14): 3059-3066. 10.1093/nar/gkf436.

Kevan P (1972) Insect pollination of High Arctic flowers. Journal of Ecology 60(3): 831-&.

Kevan P (1975) Sun-tracking solar furnaces in High Arctic flowers - Significance for pollination and insects. Science 189(4204): 723-726.

Kholina A, Koren O, Zhuravlev Y (2004) High polymorphism and autotetraploid origin of the rare endemic species *Oxytropis chankaensis* Jurtz. (Fabaceae) inferred from allozyme data. Russian Journal of Genetics 40(4): 393-400.

Kholina A, Koren' O, Zhuravlev Y (2009) Genetic structure and differentiation of populations of the tetraploid species *Oxytropis chankaensis* (Fabaceae). Russian Journal of Genetics 45(1): 70-80. 10.1134/S1022795409010104.

Kim K, Park S, Chung Y, Chung C, Kim J, Lee J (2004) Molecular cloning of low-temperature-inducible ribosomal proteins from soybean. Journal of Experimental Botany 55(399): 1153-1155.

King M, Wilson A (1975) Evolution at two levels in humans and chimpanzees. Science 188(4184): 107-116.

Klebesadel LJ (1993) Winter survival of grasses and legumes in subarctic Alaska as related to latitudinal adaptation, pre-winter storage of food reserves, and dry-matter concentration in overwintering tissues. In: University of Alaska, Agricultural and Forestry Experiment Station Bulletin. University of Alaska Fairbanks, Fairbanks, Alaska, pp i-21

Kliebenstein DJ, West MAL, van Leeuwen H, Kim K, Doerge RW, Michelmore RW, St Clair DA (2006) Genomic survey of gene expression diversity in *Arabidopsis thaliana*. Genetics 172(2): 1179-1189. 10.1534/genetics.105.049353.

Knight C, Vogel H, Kroymann J, Shumate A, Witsenboer H, Mitchell-Olds T (2006) Expression profiling and local adaptation of *Boechera holboellii* populations for

water use efficiency across a naturally occurring water stress gradient. Molecular Ecology 15(5): 1229-1237.

Koistinen KM, Soininen P, Venalainen TA, Hayrinen J, Laatikainen R, Perakyla M, Tervahauta AI, Karenlampi SO (2005) Birch PR-10c interacts with several biologically important ligands. Phytochemistry 66(21): 2524-2533. 10.1016/j.phytochem.2005.09.007.

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23(10): 1289-1291.

Kosakovsky Pond SL, Frost SDW (2005a) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21(10): 2531-2533. 10.1093/bioinformatics/bti320.

Kosakovsky Pond SL, Frost SDW (2005b) Not so different after all: A comparison of methods for detecting amino acid sites under selection. Molecular Biology and Evolution 22(5): 1208-1222. 10.1093/molbev/msi105.

Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21(5): 676-679. 10.1093/bioinformatics/bti079.

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) Automated phylogenetic detection of recombination using a genetic algorithm. Molecular Biology and Evolution 23(10): 1891-1901. 10.1093/molbev/msl051.

Kosova K, Vitamvas P, Prasil IT (2007) The role of dehydrins in plant response to cold. Biologia Plantarum 51(4): 601-617. DOI 10.1007/s10535-007-0133-6.

Kudo G, Harder L (2005) Floral and inflorescence effects on variation in pollen removal and seed production among six legume species. Functional Ecology 19(2): 245-254. 10.1111/j.1365-2435.2005.00961.x.

Kulshreshtha S, Creamer R, Sterling TM (2004) Phylogenetic relationships among New Mexico *Astragalus mollissimus* varieties and *Oxytropis* species by restriction fragment analysis. Weed Science 52(6): 984-988.

Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Briefings in Bioinformatics 9(4): 299-306. 10.1093/bib/bbn017.

Kytoviita M (2005) Asymmetric symbiont adaptation to arctic conditions could explain why High Arctic plants are non-mycorrhizal. FEMS Microbiology Ecology 53(1): 27-32.

Lai Z, Gross B, Zou Y, Andrews J, Rieseberg L (2006) Microarray analysis reveals differential gene expression in hybrid sunflower species. Molecular Ecology 15(5): 1213-1227.

Lai Z, Kane N, Zou Y, Rieseberg L (2008) Natural variation in gene expression between wild and weedy populations of *Helianthus annuus*. Genetics 179(4): 1881-1890. 10.1534/genetics.108.091041.

Langran X, Xiangyun Z, Bojian B, Mingli Z, Hang S, Podlech D, Welsh SL, Ohashi H, Larsen K, Brach AR (2010) Fabaceae: Tribe Galegeae. In: Qiner Y, Turland NJ (eds) Flora of China. Flora of China (@efloras.org)

Lauder GV, Leroi AM, Rose MR (1993) Adaptations and history. Trends in Ecology & Evolution 8(8): 294-297. 10.1016/0169-5347(93)90258-q.

Lavin M, Herendeen P, Wojciechowski M (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Systematics Biology 54(4): 575-594. doi: 10.1080/10635150590947131.

Leakey ADB, Ainsworth EA, Bernard SM, Markelz RJC, Ort DR, Placella SA, Rogers A, Smith MD, Sudderth EA, Weston DJ, Wullschleger SD, Yuan SH (2009) Gene expression profiling: opening the black box of plant ecosystem responses to global change. Global Change Biology 15(5): 1201-1213. 10.1111/j.1365-2486.2008.01818.x.

Lebel S, Schellenbaum P, Walter B, Maillot P (2010) Characterisation of the *Vitis vinifera* PR10 multigene family. BMC Plant Biology 10: 184. 10.1186/1471-2229-10-184.

Ledingham G, Rever B (1963) Chromosome numbers of some Southwest Asian species of *Astragalus* and *Oxytropis* (Leguminosae). Canadian Journal of Genetics and Cytology 5(1): 18-32.

Ledingham GF (1957) Chromosome numbers of some Saskatchewan Leguminosae with particular reference to *Astragalus* and *Oxytropis*. Canadian Journal of Botany 35(5): 657-666.

Ledingham GF (1960) Chromosome numbers in *Astragalus* and *Oxytropis*. Canadian Journal of Genetics and Cytology 2(2): 119-128. doi:10.1139/g60-012.

Liu J, Ekramoddoullah A (2006) The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses. Physiological and Molecular Plant Pathology 68(1-3): 3-13.

Lutz C (2010) Cell physiology of plants growing in cold environments. Protoplasma 244(1-4): 53-73. 10.1007/s00709-010-0161-5.

Lytle B, Song J, de la Cruz N, Peterson F, Johnson K, Bingman C, Phillips G, Volkman B (2009) Structures of two *Arabidopsis thaliana* major latex proteins represent novel helix-grip folds. Proteins - Structure Function and Bioinformatics 76(1): 237-243. 10.1002/prot.22396.

Maddison WP, Maddison DR (2011) Mesquite: a modular system for evolutionary analysis. In, 2.75 edn

Maglott D, Ostell J, Pruitt K, Tatusova T (2005) Entrez Gene: gene-centered information at NCBI. Nucleic Acids Research 33: D54-D58.

Malyshev LI (2008) Phenetics of the subgenera and sections in the genus *Oxytropis* DC. (Fabaceae) bearing on ecology and phylogeny. Contemporary Problems of Ecology 1(4): 440-444. 10.1134/s1995425508040073.

Malyshev LI (2009) Ecological enclave of the Asian genus *Oxytropis* DC (Fabaceae) in North America. Plant Life of Asian Russia 1(3): 31-43.

Marchand F, Kockelbergh F, van de Vijver B, Beyens L, Nijs I (2006) Are heat and cold resistance of arctic species affected by successive extreme temperature events? New Phytologist 170(2): 291-300.

Markovic-Housley Z, Degano M, Lamba D, von Roepenack-Lahaye E, Clemens S, Susani M, Ferreira F, Scheiner O, Breiteneder H (2003) Crystal structure of a hypoallergenic isoform of the major birch pollen allergen Bet v 1 and its likely biological function as a plant steroid carrier. Journal of Molecular Biology 325(1): 123-133.

Martienssen RA, Doerge R, Colot V (2005) Epigenomic mapping in *Arabidopsis* using tiling microarrays. Chromosome Research 13: 299-308. 10.1007/s10577-005-1507-2.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P (2010) RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26(19): 2462-2463. 10.1093/bioinformatics/btq467.

Masters BC, Fan V, Ross HA (2011) Species delimitation - a geneious plugin for the exploration of species boundaries. Molecular Ecology Resources 11(1): 154-157. 10.1111/j.1755-0998.2010.02896.x.

McCubbin A, Lee C, Hetrick A (2006) Identification of genes showing differential expression between morphs in developing flowers of *Primula vulgaris*. Sexual Plant Reproduction 19(2): 63-72.

McGuire AD, Chapin FS, Walsh JE, Wirth C (2006) Integrated regional changes in arctic climate feedbacks: Implications for the global climate system. Annual Review of Environment and Resources 31: 61-91. 10.1146/annurev.energy.31.020105.100253.

McIntosh K, Bonham-Smith P (2006) Ribosomal protein gene regulation: what about plants? Canadian Journal of Botany - Revue Canadienne de Botanique 84(3): 342-362.

Meshi T, Taoka K, Iwabuchi M (2000) Regulation of histone gene expression during the cell cycle. Plant Molecular Biology 43(5-6): 643-657.

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Research 8(11): 1113-1130.

Miroslavov E, Voznesenskaya E, Bubolo L (1996) Chloroplast structure in northern plants in relation to chloroplast adaptation to arctic conditions. Russian Journal of Plant Physiology 43(3): 325-330.

Miroslavov EA, Bubolo LS (1980) The ultrastructure of the cells of leaf chlorenchyma in some Arctic plants. Botanical Journal USSR 65(11): 1523-1530.

Miroslavov EA, Kravkina IM (1991) Comparative analysis of chloroplasts and mitochondria in leaf chlorenchyma from mountain plants grown at different altitudes. Annals of Botany 68(3): 195-200.

Mogensen JE, Wimmer R, Larsen JN, Spangfort MD, Otzen DE (2002) The major birch allergen, Bet v 1, shows affinity for a broad spectrum of physiological ligands. Journal of Biological Chemistry 277(26): 23684-23692. 10.1074/jbc.M202065200.

Mondragon-Palomino M, Hiese L, Harter A, Koch MA, Theissen G (2009) Positive selection and ancient duplications in the evolution of class B floral homeotic genes of orchids and grasses. BMC Evolutionary Biology 9. 10.1186/1471-2148-9-81.

Monroy AF, Castonguay Y, Laberge S, Sarhan F, Vezina LP, Dhindsa RS (1993) A new cold-induced alfalfa gene is associated with enhanced hardening at subzero temperature. Plant Physiology 102(3): 873-879.

Montalvo-Hernandez L, Piedra-Ibarra E, Gomez-Silva L, Lira-Carmona R, Acosta-Gallegos JA, Vazquez-Medrano J, Xoconostle-Cazares B, Ruiz-Medrano R (2008) Differential accumulation of mRNAs in drought-tolerant and susceptible common bean cultivars in response to water deficit. New Phytologist 177(1): 102-113. 10.1111/j.1469-8137.2007.02247.x.

Morinaga S, Nagano A, Miyazaki S, Kubo M, Fukuda T, Sakai S, Hasebe M, Fukuda H (2008) Ecogenomics of cleistogamous and chasmogamous flowering: genome-wide gene expression patterns from cross-species microarray analysis in *Cardamine kokaiensis* (Brassicaceae). Journal of Ecology 96(5): 1086-1097.

Mouillon JM, Eriksson SK, Harryson P (2008) Mimicking the plant cell interior under water stress by macromolecular crowding: disordered dehydrin proteins are highly resistant to structural collapse. Plant Physiology 148(4): 1925-1937. 10.1104/pp.108.124099.

Mulder CPH, Harmsen R (1995) The effect of muskox herbivory on growth and reproduction in an arctic legume. Arctic and Alpine Research 27(1): 44-53.

Murray C, Miller P (1982) Phenological observations of major plant-growth forms and species in montane and *Eriophorum-vaginatum* tussock tundra in Central Alaska. Holarctic Ecology 5(2): 109-116.

Musidlowska-Persson A, Alm R, Emanuelsson C (2007) Cloning and sequencing of the Bet v 1-homologous allergen Fra a 1 in strawberry (*Fragaria ananassa*) shows the

presence of an intron and little variability in amino acid sequence. Molecular Immunology 44(6): 1245-1252. 10.1016/j.molimm.2006.06.004.

National Climate Data and Information Archive (2009) Canadian climate normals or averages 1971-2000. Environment Canada. http://www.climat.meteo.gc.ca/climate_normals/index_e.html?&. Accessed 2009.

Natural Resources Canada (1995a) The Atlas of Canada; Permafrost. http://atlas.nrcan.gc.ca/site/english/maps/environment/land/permafrost.

Natural Resources Canada (1995b) The Atlas of Canada: July mean daily minimum and maximum temperatures. http://atlas.nrcan.gc.ca/auth/english/maps/environment/climate/temperature/temp_summer.

Natural Resources Canada (1995c) The Atlas of Canada: Land Cover. http://atlas.nrcan.gc.ca/site/english/maps/environment/land/landcover/1.

Natural Resources Canada (1995d) The Atlas of Canada: Terrestrial Ecozones. http://atlas.nrcan.gc.ca/site/english/maps/environment/forest/forestcanada/terrestrialecozones/interactivemap_view?scale=42051275.911682&mapsize=1150%201350&map_web=TEMPLATE%20/home/atlas/mapdata/ajax.html&urlappend=.

Nielsen R (2005) Molecular signatures of natural selection. Annual Review of Genetics 39: 197-218. 10.1146/annurev.genet.39.073003.112420.

Nylander J (2004) MrModeltest v2. Program distributed by the author. In, v2 edn. Evolutionary Biology Centre, Uppsala University

Oh DH, Dassanayake M, Bohnert HJ, Cheeseman JM (2012) Life at the extreme: lessons from the genome. Genome Biology 13(3). 10.1186/gb-2012-13-3-241.

Osmark P, Boyle B, Brisson N (1998) Sequential and structural homology between intracellular pathogenesis-related proteins and a group of latex proteins. Plant Molecular Biology 38: 1243-1246.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20(2): 289-290. 10.1093/bioinformatics/btg412.

Pasternak O, Biesiadka J, Dolot R, Handschuh L, Bujacz G, Sikorski M, Jaskolski M (2005) Structure of a yellow lupin pathogenesis-related PR-10 protein belonging to a novel

subclass. Acta Crystallographica Section D(Biological Crystallography): 99-107. 10.1107/S0907444904028173.

Pasternak O, Bujacz G, Fujimoto Y, Hashimoto Y, Jelen F, Otlewski J, Sikorski M, Jaskolski M (2006) Crystal structure of *Vigna radiata* cytokinin-specific binding protein in complex with zeatin. Plant Cell 18: 2622-2634. 10.1105/tpc.105.037119.

Pennycooke J, Cheng H, Stockinger E (2008) Comparative genomic sequence and expression analyses of *Medicago truncatula* and alfalfa subspecies *falcata* COLD-ACCLIMATION-SPECIFIC genes. Plant Physiology 146(3): 1242-1254. 10.1104/pp.107.108779.

Pinto M, Ribeiro A, Regalado A, Rodrigues-Pousada C, Ricardo C (2005) Expression of *Lupinus albus* PR-10 proteins during root and leaf development. Biologia Plantarum 49(2): 187-193.

Pinto M, Ricardo C (1995) *Lupinus albus* L pathogenesis-related proteins that show similarity to PR-10 proteins. Plant Physiology 109(4): 1345-1351.

Plessl M, Rigola D, Hassinen VH, Tervahauta A, Karenlampi S, Schat H, Aarts MGM, Ernst D (2010) Comparison of two ecotypes of the metal hyperaccumulator *Thlaspi caerulescens* (J. & C. Presl) at the transcriptional level. Protoplasma 239(1-4): 81-93. 10.1007/s00709-009-0085-0.

Prevost D, Bordeleau L, Caudryreznick S, Schulman H, Antoun H (1987) Characteristics of rhizobia isolated from three legumes indigenous to the Canadian High Arctic: *Astragalus alpinus*, *Oxytropis maydelliana*, and *Oxytropis arctobia*. Plant and Soil 98(3): 313-324.

Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, Van de Peer Y, Vandepoele K (2009) PLAZA: A comparative genomics resource to study gene and genome evolution in plants. Plant Cell 21(12): 3718-3731. 10.1105/tpc.109.071506.

Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. Trends in Ecology & Evolution 23(10): 564-571. 10.1016/j.tree.2008.06.010.

Pryor B, Creamer R, Shoemaker R, McLain-Romero J, Hambleton S (2009) *Undifilum*, a new genus for endophytic *Embellisia oxytropis* and parasitic *Helminthosporium bornmuelleri* on legumes. Botany-Botanique 87(2): 178-194. DOI 10.1139/B08-130.

Puckette M, Peal L, Steele J, Tang YH, Mahalingam R (2009) Ozone responsive genes in *Medicago truncatula*: Analysis by suppression subtraction hybridization. Journal of Plant Physiology 166(12): 1284-1295. 10.1016/j.jplph.2009.01.009.

Puckette MC, Tang YH, Mahalingam R (2008) Transcriptomic changes induced by acute ozone in resistant and sensitive *Medicago truncatula* accessions. BMC Plant Biology 8. 10.1186/1471-2229-8-46.

Pyankov V (1991) Effect of temperature on photosynthesis and carbon metabolism in related plant species of the arctic tundra of Wrangel Island and the central Urals. Soviet Plant Physiology 38(5): 607-613.

Pyankov V, Vaskovskii M (1994) Temperature adaptation of the photosynthetic apparatus of tundra plants *Oxyria digyna* and *Alopecurus alpinus* from Wrangel Island. Russian Journal of Plant Physiology 41(4): 454-461.

Qian G, Liu YH, Ao DS, Yang F, Yu MQ (2008) Differential expression of dehydrin genes in hull-less barley (*Hordeum vulgare* ssp *vulgare*) depending on duration of dehydration stress. Canadian Journal of Plant Science 88(5): 899-906.

Quesada T, Li Z, Dervinis C, Li Y, Bocock PN, Tuskan GA, Casella G, Davis JM, Kirst M (2008) Comparative analysis of the transcriptomes of *Populus trichocarpa* and *Arabidopsis thaliana* suggests extensive evolution of gene expression regulation in angiosperms. New Phytologist 180(2): 408-420. 10.1111/j.1469-8137.2008.02586.x.

Rahman LN, Chen L, Nazim S, Bamm VV, Yaish MW, Moffatt BA, Dutcher JR, Harauz G (2010) Interactions of intrinsically disordered *Thellungiella salsuginea* dehydrins TsDHN-1 and TsDHN-2 with membranes - synergistic effects of lipid composition and temperature on secondary structure. Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire 88(5): 791-807. 10.1139/o10-026.

Ralphs M, Creamer R, Baucom D, Gardner D, Welsh S, Graham J, Hart C, Cook D, Stegelmeier B (2008) Relationship between the endophyte *Embellisia* spp. and the toxic alkaloid swainsonine in major locoweed species (*Astragalus* and *Oxytropis*). Journal of Chemical Ecology 34(1): 32-38. 10.1007/s10886-007-9399-6.

Remus-Borel W, Castonguay Y, Cloutier J, Michaud R, Bertrand A, Desgagnes R, Laberge S (2010) Dehydrin variants associated with superior freezing tolerance in alfalfa

(*Medicago sativa* L.). Theoretical and Applied Genetics 120(6): 1163-1174. 10.1007/s00122-009-1243-7.

Rinne PLH, Kaikuranta PLM, van der Plas LHW, van der Schoot C (1999) Dehydrins in cold-acclimated apices of birch (*Betula pubescens* Ehrh.): production, localization and potential role in rescuing enzyme function during dehydration. Planta 209(4): 377-388.

Ritz C, Spiess A-N (2008) qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. Bioinformatics 24(13): 1549-1551. 10.1093/bioinformatics/btn227.

Rizzon C, Ponger L, Gaut BS (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. Plos Computational Biology 2(9): 989-1000. 10.1371/journal.pcbi.0020115.

Robberecht R, Junttila O (1992) The freezing response of an arctic cushion plant, *Saxifraga caespitosa* L - acclimation, freezing tolerance and ice nucleation. Annals of Botany 70(2): 129-135.

Rorat T (2006) Plant dehydrins - Tissue location, structure and function. Cellular & Molecular Biology Letters 11(4): 536-556. DOI 10.2478/s11658-006-0044-0.

Rozen S, Skaletsky H (2000) Primer 3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols. The Humana Press Inc., Totowa, New Jersey, pp 365–386

Ruepp A, Zollner A, Maier D, Albermann K, Hani J, Mokrejs M, Tetko I, Guldener U, Mannhaupt G, Munsterkotter M, Mewes H (2004) The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucleic Acids Research 32(18): 5539-5545.

Russell R, Cutler D, Jacobs S, King A, Pollard A (1940) Physiological and ecological studies on an arctic vegetation II The development of vegetation in relation to nitrogen supply and soil micro organisms on Jan Mayen island. Journal of Ecology 28(1): 269-288.

Saez-Vasquez J, Gallois P, Delseny M (2000) Accumulation and nuclear targeting of BnC24, a *Brassica napus* ribosomal protein corresponding to a mRNA accumulating in response to cold treatment. Plant Science 156(1): 35-46.

Savile D (1972) Arctic adaptations in plants. Canada Dept. of Agriculture, Ottawa

Savitch L, Barker-Astrom J, Ivanov A, Hurry V, Oquist G, Huner N, Gardestrom P (2001) Cold acclimation of *Arabidopsis thaliana* results in incomplete recovery of photosynthetic capacity, associated with an increased reduction of the chloroplast stroma. Planta 214(2): 295-303.

Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Mizrachi I, Ostell J, Panchenko A, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, John Wilbur W, Yaschenko E, Ye J (2010) Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 38(supplement 1 Database issue): D5-16. 10.1093/nar/gkp967.

Schenk MF, Gilissen L, Esselink GD, Smulders MJM (2006) Seven different genes encode a diverse mixture of isoforms of Bet v I, the major birch pollen allergen. BMC Genomics 7. 10.1186/1471-2164-7-168.

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. Nature Genetics 37(5): 501-506. 10.1038/ng1543.

Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu SQ, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du JC, Tian ZX, Zhu LC, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463(7278): 178-183. Doi 10.1038/Nature08670.

Schonswetter P, Tribsch A, Niklfeld H (2004) Amplified fragment length polymorphism (AFLP) reveals no genetic divergence of the Eastern Alpine endemic *Oxytropis campestris* subsp *tiroliensis* (Fabaceae) from widespread subsp campestris. Plant Systematics and Evolution 244: 245-255. 10.1007/s00606-003-0096-9.

Semikhatova O, Ivanova T, Kirpichnikova O (2007) Comparative study of dark respiration in plants inhabiting arctic (Wrangel Island) and temperate climate zones. Russian Journal of Plant Physiology 54(5): 582-588.

Shoemaker R, Keim P, Vodkin L, Retzel E, Clifton S, Waterston R, Smoller D, Coryell V, Khanna A, Erpelding J, Gai X, Brendel V, Raph-Schmidt C, Shoop E, Vielweber C, Schmatz M, Pape D, Bowers Y, Theising B, Martin J, Dante M, Wylie T, Granger C (2002) A compilation of soybean ESTs: generation and analysis. Genome 45(2): 329-338.

Sikorski M, Biesiadka J, Kasperska A, Kopcinska J, Lotocka B, Golinowski W, Legocki A (1999) Expression of genes encoding PR10 class pathogenesis-related proteins is inhibited in yellow lupine root nodules. Plant Science 149(2): 125-137.

Simon P (2003) Q-Gene: processing quantitative real-time RT-PCR data. Bioinformatics 19(11): 1439-1440.

Siol M, Wright SI, Barrett SCH (2010) The population genomics of plant adaptation. New Phytologist 188(2): 313-332. 10.1111/j.1469-8137.2010.03401.x.

Siomos MF (2009) Shaped by the environment - adaptation in plants. FEBS Journal 276(17): 4705-4714. 10.1111/j.1742-4658.2009.07170.x.

Skrede I, Borgen L, Brochmann C (2009) Genetic structuring in three closely related circumpolar plant species: AFLP versus microsatellite markers and high-arctic versus arctic-alpine distributions. Heredity 102(3): 293-302. 10.1038/hdy.2008.120.

Slotte T, Holm K, McIntyre LM, Lagercrantz U, Lascoux M (2007) Differential expression of genes important for adaptation in *Capsella bursa-pastoris* (Brassicaceae). Plant Physiology 145(1): 160-173. 10.1104/pp.107.102632.

Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (2007) Climate Change 2007 the physical science basis: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on

Climate Change, 2007. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Sørensen T (1941) Temperature relations and phenology of the northeast Greenland flowering plants. Meddelelser om Grønland 125(9): 1-305.

Srivastava S, Emery RJN, Kurepin LV, Reid DM, Fristensky B, Kav NNV (2006) Pea PR 10.1 is a ribonuclease and its transgenic expression elevates cytokinin levels. Plant Growth Regulation 49(1): 17-25. 10.1007/s10725-006-0022-6.

Stegelmeier B, James L, Panter K, Gardner D, Pfister J, Ralphs M, Molyneux R (1999) Dose response of sheep poisoned with locoweed (*Oxytropis sericea*). Journal of Veterinary Diagnostic Investigation 11(5): 448-456.

Stitt M, Hurry V (2002) A plant for all seasons: alterations in photosynthetic carbon metabolism during cold acclimation in *Arabidopsis*. Current Opinion in Plant Biology 5(3): 199-206.

Storz J (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. Molecular Ecology 14: 671-688. DOI 10.1111/j.1365-294X.2004.02437.x.

Strathdee A, Bale J (1998) Life on the edge: Insect ecology in arctic environments. Annual Review of Entomology 43: 85-106.

Strömvik M, Latour F, Archambault A, Vodkin L (2006) Identification and phylogenetic analysis of sequences of *Bean pod mottle virus*, *Soybean mosaic virus*, and *Cowpea chlorotic mottle virus* in expressed sequence tag data from soybean. Canadian Journal of Plant Pathology - Revue Canadienne de Phytopathologie 28(2): 289-301.

Strömvik M, Sundararaman V, Vodkin L (1999) A novel promoter from soybean that is active in a complex developmental pattern with and without its proximal 650 base pairs. Plant Molecular Biology 41(2): 217-231.

Sun Y, Skinner DZ, Liang GH, Hulbert SH (1994) Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. Theoretical and Applied Genetics 89(1): 26-32.

Svoboda J (2009) Evolution of plant cold hardiness and its manifestation along the latitudinal gradient in the Canadian Arctic. In: Gusta LV, Tanino KK, Wisniewski

ME (eds) Plant cold hardiness : from the laboratory to the field. CABI, Wallingford Oxfordshire, UK; Cambridge, MA, pp 140-162

Swarbreck D, Wilks C, Lamesch P, Berardini T, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Research 36: D1009-D1014.

Swindell W, Huebner M, Weber A (2007) Plastic and adaptive gene expression patterns associated with temperature stress in *Arabidopsis thaliana*. Heredity 99(2): 143-150.

Swofford D (2002) PAUP*: Phylogenetic analysis using parsimony (*and other methods). In. Sinauer, Sunderland, Massachusetts, USA

Taji T, Sakurai T, Mochida K, Ishiwata A, Kurotani A, Totoki Y, Toyoda A, Sakaki Y, Seki M, Ono H, Sakata Y, Tanaka S, Shinozaki K (2008) Large-scale collection and annotation of full-length enriched cDNAs from a model halophyte, *Thellungiella halophila*. BMC Plant Biology 8(1): 115.

Taji T, Seki M, Satou M, Sakurai T, Kobayashi M, Ishiyama K, Narusaka Y, Narusaka M, Zhu J, Shinozaki K (2004) Comparative genomics in salt tolerance between *Arabidopsis* and *Arabidopsis*-related halophyte salt cress using *Arabidopsis* microarray. Plant Physiology 135(3): 1697-1709. 10.1104/pp.104.039909.

Takahashi R, Shimosaka E (1997) cDNA sequence analysis and expression of two cold-regulated genes in soybean. Plant Science 123(1-2): 93-104.

Teeri JA (1976) Phytotron analysis of a photoperiodic response in a High Arctic plant species. Ecology 57(2): 374-379.

Tenaillon MI, Tiffin PL (2008) The quest for adaptive evolution: a theoretical challenge in a maze of data. Current Opinion in Plant Biology 11(2): 110-115. 10.1016/j.pbi.2007.12.003.

Terras FRG (1995) Small cysteine-rich antifungal proteins from radish: their role in host defense. Plant Cell 7(5): 573-588.

Thomashow MF (2010) Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway. Plant Physiology 154(2): 571-577. 10.1104/pp.110.161794.

Thomma B, Cammue B, Thevissen K (2002) Plant defensins. Planta 216(2): 193-202.

Tirosh I, Weinberger A, Carmi M, Barkai N (2006) A genetic signature of interspecies variations in gene expression. Nature Genetics 38(7): 830-834. 10.1038/ng1819.

Toleno DM, Durbin ML, Lundy KE, Clegg MT (2010) Extensive evolutionary rate variation in floral color determining genes in the genus *Ipomoea*. Plant Species Biology 25(1): 30-42. 10.1111/j.1442-1984.2009.00256.x.

Totland O, Alatalo J (2002) Effects of temperature and date of snowmelt on growth, reproduction, and flowering phenology in the arctic/alpine herb, *Ranunculus glacialis*. Oecologia 133(2): 168-175.

Travers SE, Tang ZW, Caragea D, Garrett KA, Hulbert SH, Leach JE, Bai JF, Saleh A, Knapp AK, Fay PA, Nippert J, Schnable PS, Smith MD (2010) Variation in gene expression of *Andropogon gerardii* in response to altered environmental conditions associated with climate change. Journal of Ecology 98(2): 374-383. 10.1111/j.1365-2745.2009.01618.x.

Valdez Barillas J, Paschke M, Ralphs M, Child R (2007) White locoweed toxicity is facilitated by a fungal endophyte and nitrogen-fixing bacteria. Ecology 88(7): 1850-1856.

van de Mortel J, Villanueva L, Schat H, Kwekkeboom J, Coughlan S, Moerland P, van Themaat E, Koornneef M, Aarts M (2006) Large expression differences in genes for iron and zinc homeostasis, stress response, and lignin biosynthesis distinguish roots of *Arabidopsis thaliana* and the related metal hyperaccumulator *Thlaspi caerulescens*. Plant Physiology 142(3): 1127-1147.

van der Aa LM, Levraud JP, Yahmi M, Lauret E, Briolat V, Herbomel P, Benmansour A, Boudinot P (2009) A large new subset of TRIM genes highly diversified by duplication and positive selection in teleost fish. BMC Biology 7. 10.1186/1741-7007-7-7.

van Leeuwen H, Kliebenstein DJ, West MAL, Kim K, van Poecke R, Katagiri F, Michelmore RW, Doerge RW, Clair DA (2007) Natural variation among *Arabidopsis thaliana* accessions for transcriptome response to exogenous salicylic acid. Plant Cell 19(7): 2099-2110. 10.1105/tpc.107.050641.

Vaughn M, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz P, Dedhia N, McCombie W, Agier N, Bulski A, Colot V, Doerge R, Martienssen R (2007) Epigenetic natural variation in *Arabidopsis thaliana*. Plos Biology 5(7): 1617-1629. ARTN e174.

Vodkin L, Khanna A, Shealy R, Clough S, Gonzalez D, Philip R, Zabala G, Thibaud-Nissen F, Sidarous M, Strömvik M, Shoop E, Schmidt C, Retzel E, Erpelding J, Shoemaker R, Rodriguez-Huete A, Polacco J, Coryell V, Keim P, Gong G, Liu L, Pardinas J, Schweitzer P (2004) Microarrays for global expression constructed with a low redundancy set of 27,500 sequenced cDNAs representing an array of developmental stages and physiological conditions of the soybean plant. BMC Genomics 5: 73. DOI 10.1186/1471-2164-5-73.

Voelckel C, Heenan P, Janssen B, Reichelt M, Ford K, Hofmann R, Lockhart P (2008) Transcriptional and biochemical signatures of divergence in natural populations of two species of New Zealand alpine *Pachycladon*. Molecular Ecology 17(21): 4740-4753.

von Meijenfeldt N (2010) Unraveling the cold response in *Draba*. In: Institute for Biodiversity and Ecosystem Dynamics. University of Amsterdam, Nederlands, p 147

Vuylsteke M, Van Den Daele H, Vercauteren A, Zabeau M, Kuiper M (2006) Genetic dissection of transcriptional regulation by cDNA-AFLP. Plant Journal 45(3): 439-446.

Wager H (1938) Growth and survival of plants in the arctic. Journal of Ecology 26(2): 390-410.

Walker M, Wahren C, Hollister R, Henry G, Ahlquist L, Alatalo J, Bret-Harte M, Calef M, Callaghan T, Carroll A, Epstein H, Jonsdottir I, Klein J, Magnusson B, Molau U, Oberbauer S, Rewa S, Robinson C, Shaver G, Suding K, Thompson C, Tolvanen A, Totland O, Turner P, Tweedie C, Webber P, Wookey P (2006) Plant community responses to experimental warming across the tundra biome. Proceedings of the National Academy of Sciences of the United States of America 103(5): 1342-1346.

Walters R (2005) Towards an understanding of photosynthetic acclimation. Journal of Experimental Botany 56(411): 435-447. DOI 10.1093/jxb/eri060.

Wang HG, Zhang HL, Gao FH, Li JX, Li ZC (2007) Comparison of gene expression between upland and lowland rice cultivars under water stress using cDNA microarray. Theoretical and Applied Genetics 115(8): 1109-1126. 10.1007/s00122-007-0637-7.

Wang J, Tian L, Lee H, Wei N, Jiang H, Watson B, Madlung A, Osborn T, Doerge R, Comai L, Chen Z (2006) Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. Genetics 172(1): 507-517.

Wang WX, Wei LB (2003) Purification of boiling-soluble antifreeze protein from the legume *Ammopiptanthus mongolicus*. Preparative Biochemistry & Biotechnology 33(1): 67-80. 10.1081/pb-120018370.

Wang X, He R, He G (2005) Construction of suppression subtractive hybridization libraries and identification of brown planthopper-induced genes. Journal of Plant Physiology 162(11): 1254-1262.

Warren AS, Anandakrishnan R, Zhang LQ (2010) Functional bias in molecular evolution rate of *Arabidopsis thaliana*. BMC Evolutionary Biology 10. 10.1186/1471-2148-10-125.

Welsh SL (1991) *Oxytropis* DC., names, basionyms, types, and synonyms: Flora of North America project. Great Basin Naturalist 51(4): 377-396.

Welsh SL (2001) Revision of North American species of *Oxytropis* de Candolle (Leguminosae). EPS Inc., Orem, Utah

Whitehead A, Crawford D (2006a) Neutral and adaptive variation in gene expression. Proceedings of the National Academy of Sciences of the United States of America 103(14): 5425-5430.

Whitehead A, Crawford D (2006b) Variation within and among species in gene expression: raw material for evolution. Molecular Ecology 15(5): 1197-1211.

Whittle C, Krochko J (2009) Transcript profiling provides evidence of functional divergence and expression networks among ribosomal protein gene paralogs in *Brassica napus*. Plant Cell 21(8): 2203-2219. 10.1105/tpc.109.068411.

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ (2007) An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. Plos One 2(8). 10.1371/journal.pone.0000718.

Wise R, Caldo R, Hong L, Wu S, Cannon E, Dickerson J (2006) PLEXdb: A unified expression profiling database for plants and plant pathogens. Phytopathology 96(6): S161-S161.

Wojciechowski M (2005) *Astragalus* (Fabaceae): A molecular phylogenetic perspective. Brittonia: 382-396.

Wojciechowski MF, Sanderson MJ, Baldwin BG, Donoghue MJ (1993) Monophyly of aneuploid *Astragalu*s (Fabaceae) Evidence from nuclear ribosomal DNA internal transcribed spacer sequences. American Journal of Botany 80(6): 711-722.

Wojciechowski MF, Sanderson MJ, Hu JM (1999) Evidence on the monophyly of *Astragalus* (Fabaceae) and its major subgroups based on nuclear ribosomal DNA ITS and chloroplast DNA *trn*L intron data. Systematic Botany 24(3): 409-437.

Wong CE, Li Y, Labbe A, Guevara D, Nuin P, Whitty B, Diaz C, Golding GB, Gray GR, Weretilnyk EA, Griffith M, Moffatt BA (2006) Transcriptional profiling implicates novel interactions between abiotic stress and hormonal responses in *Thellungiella*, a close relative of *Arabidopsis*. Plant Physiology 140(4): 1437-1450. 10.1104/pp.105.070508.

Wong CE, Li Y, Whitty BR, Diaz-Camino C, Akhter SR, Brandle JE, Golding GB, Weretilnyk EA, Moffatt BA, Griffith M (2005) Expressed sequence tags from the Yukon ecotype of Thellungiella reveal that gene ex*pression in r*esponse to cold, drought and salinity shows little overlap. Plant Molecular Biology 58(4): 561-574. 10.1007/s11103-005-6163-6.

Wu HJ, Zhang ZH, Wang JY, Oh DH, Dassanayake M, Liu BH, Huang QF, Sun HX, Xia R, Wu YR, Wang YN, Yang Z, Liu Y, Zhang WK, Zhang HW, Chu JF, Yan CY, Fang S, Zhang JS, Wang YQ, Zhang FX, Wang GD, Lee SY, Cheeseman JM, Yang BC, Li B, Min JM, Yang LF, Wang J, Chu CC, Chen SY, Bohnert HJ, Zhu JK, Wang XJ, Xie Q (2012) Insights into salt tolerance from the genome of

Thellungiella salsuginea. Proceedings of the National Academy of Sciences of the United States of America 109(30): 12219-12224. 10.1073/pnas.1209954109.

Xiong F, Ruhland C, Day T (1999) Photosynthetic temperature response of the Antarctic vascular plants *Colobanthus quitensis* and *Deschampsia antarctica*. Physiologia Plantarum 106(3): 276-286.

Xu GX, Ma H, Nei M, Kong HZ (2009) Evolution of F-box genes in plants: Different modes of sequence divergence and their relationships with functional diversification. Proceedings of the National Academy of Sciences of the United States of America 106(3): 835-840. 10.1073/pnas.0812043106.

Xu JC, Belanger F, Huang BR (2008) Differential gene expression in shoots and roots under heat stress for a geothermal and non-thermal *Agrostis* grass species contrasting in heat tolerance. Environmental and Experimental Botany 63(1-3): 240-247. 10.1016/j.envexpbot.2007.11.011.

Yakovlev GP, Sytin AK, Roskov YR (1996) Legumes of Northern Eurasia: a checklist. Royal Botanic Gardens, Richmond, UK

Yamamoto YY, Obokata J (2008) ppdb: a plant promoter database. Nucleic Acids Research 36: D977-D981. 10.1093/nar/gkm785.

Yan Q, Qi X, Jiang Z, Yang S, Han L (2008) Characterization of a pathogenesis-related class 10 protein (PR-10) from *Astragalus mongholicus* with ribonuclease activity. Plant Physiology and Biochemistry 46(1): 93-99.

Yang S, Loopstra C (2005) Seasonal variation in gene expression for loblolly pines (*Pinus taeda*) from different geographical regions. Tree Physiology 25(8): 1063-1073.

Yang ZJ, Zhang T, Bolshoy A, Beharav A, Nevo E (2009) Adaptive microclimatic structural and expressional dehydrin 1 evolution in wild barley, *Hordeum spontaneum*, at 'Evolution Canyon', Mount Carmel, Israel. Molecular Ecology 18(9): 2063-2075. 10.1111/j.1365-294X.2009.04140.x.

Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H, Van de Peer Y, Proost S, Cook DR, Meyers BC, Spannagl M, Cheung F, De Mita S, Krishnakumar V, Gundlach H, Zhou S, Mudge J, Bharti AK, Murray JD, Naoumkina MA, Rosen B, Silverstein KAT, Tang H,

Rombauts S, Zhao PX, Zhou P, Barbe V, Bardou P, Bechner M, Bellec A, Berger A, Berges H, Bidwell S, Bisseling T, Choisne N, Couloux A, Denny R, Deshpande S, Dai X, Doyle JJ, Dudez A-M, Farmer AD, Fouteau S, Franken C, Gibelin C, Gish J, Goldstein S, Gonzalez AJ, Green PJ, Hallab A, Hartog M, Hua A, Humphray SJ, Jeong D-H, Jing Y, Jocker A, Kenton SM, Kim D-J, Klee K, Lai H, Lang C, Lin S, Macmil SL, Magdelenat G, Matthews L, McCorrison J, Monaghan EL, Mun J-H, Najar FZ, Nicholson C, Noirot C, O/'Bleness M, Paule CR, Poulain J, Prion F, Qin B, Qu C, Retzel EF, Riddle C, Sallet E, Samain S, Samson N, Sanders I, Saurat O, Scarpelli C, Schiex T, Segurens B, Severin AJ, Sherrier DJ, Shi R, Sims S, Singer SR, Sinharoy S, Sterck L, Viollet A, Wang B-B, Wang K, Wang M, Wang X, Warfsmann J, Weissenbach J, White DD, White JD, Wiley GB, Wincker P, Xing Y, Yang L, Yao Z, Ying F, Zhai J, Zhou L, Zuber A, Denarie J, Dixon RA, May GD, Schwartz DC, Rogers J, Quetier F, Town CD, Roe BA (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature 480: 520-524. doi:10.1038/nature10625.

Yurtsev B (1997) Analysis of evolutionary differentiation in some key arctic-alpine taxa: *Dryas*, *Oxytropis* sect. Arctobia and *Taraxacum* sect. Arctica. Opera Botanica 132: 27-37.

Yurtsev B (1999) Survey of arctic legumes with emphasis on the species concept in *Oxytropis*. In: Nordal I, Elven R (eds) The Species Concept in the High North - A Panarctic Flora Initiative. Norwegian Academy of Science and Letters, Oslo, Norway, pp 295-318

Zamora A, Sun Q, Hamblin M, Aquadro C, Kresovich S (2009) Positively selected disease response orthologous gene sets in the cereals identified using *Sorghum bicolor* L. Moench expression profiles and comparative genomics. Molecular Biology and Evolution 26(9): 2015-2030. DOI 10.1093/molbev/msp114.

Zhang P, Foerster H, Tissier C, Mueller L, Paley S, Karp P, Rhee S (2005a) MetaCyc and AraCyc. Metabolic pathway databases for plant research. Plant Physiology 138(1): 27-37.

Zhang X, Byrnes JK, Gal TS, Li WH, Borevitz JO (2008) Whole genome transcriptome polymorphisms in *Arabidopsis thaliana*. Genome Biology 9(11): 48. R165.

Zhang Y, Mian M, Chekhovskiy K, So S, Kupfer D, Lai H, Roe B (2005b) Differential gene expression in *Festuca* under heat stress conditions. Journal of Experimental Botany 56(413): 897-907.

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR *Arabidopsis* microarray database and analysis toolbox. Plant Physiology 136(1): 2621-2632. 10.1104/pp.104.046367.

8 List of Appendices

8.1 Appendix 1 : Supplementary Table S3.1 Seed sources of *Oxytropis* species for the library subtraction and real-time RT-PCR.

Oxytropis species	Locality	Collector	RNA extracted for library subtraction	RNA extracted for real-time RT-PCR
From the Arctic region				
O. arctobia Bunge	Baffin Island (Nunavut, Canada)	Dr. Susan Aiken, Canadian Museum of Nature, Ottawa	No	Yes
O. arctobia	Melville Peninsula (Nunavut, Canada)	Dr. Danielle Prévost, Agriculture and Agri-Food Canada, Ste-Foy (Prevost et al. 1987)	Yes	Yes
O. maydelliana Trautv.	Baffin Island (Nunavut, Canada)	Dr. Susan Aiken, Canadian Museum of Nature, Ottawa	Yes	Yes
O. maydelliana	Melville Peninsula Nunavut, Canada	Dr. Danielle Prévost, Agriculture and Agri-Food Canada, Ste-Foy (Prevost et al. 1987)	Yes	Yes
O. maydelliana	Seward Peninsula (Alaska), USA	Carolyn Parker, University of Alaska Fairbanks Museum of the North (ALA V97877)	Yes	No

From the temperate regions				
O. campestris (L.) DC subsp. *johannensis* (Fernald) Blondeau & Gervais	Ile d'Orléans, Quebec, Canada	Annie Archambault	Yes	Yes
O. splendens Douglas	Southern Alberta, Canada	Plant Gene Resources of Canada (Ottawa) (CN105143)	Yes	No
O. splendens	Longview, Southern Alberta, Canada	Wild About Flowers, native plant nursery	Yes	Yes
O. splendens	Cochin, Southern Saskatchewan, Canada	Prairie Garden Seeds, native plant nursery	No	Yes

8.2 Appendix 2: Supplementary Table S3.2 Primers used for real-time reverse transcription PCR (RT-PCR)[a] from *Oxytropis* plantlets cDNA for PR-10, KS-dehydrin and light harvesting proteins (*LHCB*).

Gene name	Forward primer (5' – 3')	Reverse primer (5' – 3')	Annealing temperature	Amplicon length	Sequences accession numbers
arctic.contig13/36 (PR-10 family)	PR10_13_1f CTCCAAGTCATCAATTCATCACCAC	PR10_36_4r TTAGGATTGGCCAAAACATAACC	63 °C	397 bp	HM107140 (Oa) HM107141 (Os) HM107142 (Ocj) HM107143 (Om)
arctic.contig61 (PR-10 family)	PR10_61_1f CAACACAACACAACWGCCAACAA	PR10_61_16r AAACGTAACCCTCGATAGCCT	63° C	524 bp	HM107135 (Oa) HM107136 (Ocj) HM107137 (Os) HM107138 (Os) HM107139 (Om)
arctic.contig47 (KS-dehydrin)	arct_cold47_7f GAGAACACTATGGTGAACCACAT	arct_cold47_8r TGCTGCTGCTATCATGACCATGC	63° C	288 bp	HM107144 (Os) HM107145 (Om) HM107146 (Ocj) HM107147 (Oa)
temperate.contig119 (lhcb1, light harvesting protein b 1)	lhcb1_119_5f GCCAAAAACCGTGAACTCGAAG	lhcb1_119_7r TGGGTCAACAACCTCACCAAGA	63° C	288 bp	HM107148 (Ocj) HM107149 (Om) HM107150 (Os) HM107151 (Oa)
actin (as normalizing gene)	actinO_7f ATGTGCCTGCCATGTATGTTGC	actinO_8r GCAAGATCCAAACGAAGGATGG	63° C	220 bp	HM107152 (Oa) HM107153 (Os) HM107154 (Om) HM107155 (Ocj)

8.3 Appendix 3: Supplementary Table S3.3 Annotation of the 121 genes (contigs and singlets) from an arctic-enriched *Oxytropis* subtracted plantlet library.

Arctic unique sequence name	Nb of ESTs in contig	Gene putative identity	Length	Number of BLAST hits	Max e-value	Similarity mean	Number of GO terms	GO terms	Enzyme	Member sequences	GenBank Accession number
		Photosynthesis photosystems									
*Oapa1_0329_CZ1835_078.ab1	1	psaB photosystem i p700 apoprotein a2; electron transport	471	25	1.36E-56	99.20%	14	F:electron carrier activity; F:iron ion binding; P:transport; P:oxidation reduction; F:magnesium ion binding; P:protein-chromophore linkage; F:chlorophyll binding; F:4 iron, 4 sulfur cluster binding; C:integral to membrane; C:chloroplast; C:thylakoid; C:photosystem I; P:photosynthesis; P:electron transport			GW697081
		proton transport									

arctic.fasta.screen.Contig37	2	atpE atp synthase cf1 epsilon subunit; photosynthesis and oxydative phosphorylation; proton transport; chloroplast protein encoded by chloroplast genome	638	25	2.27E-82	96.60%	8	C:proton-transporting ATP synthase complex, catalytic core F(1); P:plasma membrane ATP synthesis coupled proton transport; F:hydrogen ion transporting ATP synthase activity, rotational mechanism; F:metal ion binding; C:thylakoid; F:ATP binding; F:hydrogen ion transporting ATPase activity, rotational mechanism; C:chloroplast	EC:3.6.3.14,	Oapa1_0082_HFY327_056.ab1; Oapa1_0588_CZ1855_068.ab1	GW696989 GW697178
Oapa1_0421_CZ1835_040.ab1	1	Carbohydrates other carbohydrate metabolism Acidic endochitinase precursor, aminosugar metabolism Nitrogen metabolism	709	25	9.15E-89	80.80%	5	F:chitinase activity; P:chitin catabolic process; F:cation binding; C:extracellular region; P:cell wall catabolic process	EC:3.2.1.14,		GW697122

*Oapa1_1465_AM23_027.ab1	1	- Amino acids Carbonic anhydrase 1 (ca1) carbonate dehydratase zinc ion binding Protein folding and modifications	315	3	7.70E-07	62.00%	5	C:stromule; C:apoplast; C:thylakoid; F:protein binding; C:chloroplast		GW697525
Oapa1_1334_EY0016_001.ab1	1	Calreticulin precursor ; Chaperones and folding catalysts; Glycan Binding	712	25	4.88E-70	94.88%	10	F:sugar binding; P:protein folding; C:endoplasmic reticulum lumen; P:response to oxidative stress; P:response to cadmium ion; C:plasma membrane; F:calcium ion binding; F:unfolded protein binding; F:zinc ion binding; C:mitochondrion		GW697469
*Oapa1_0946_LS228_065.ab1	1	Signaling calcium related Calmodulin ; Phosphatidylinositol signaling	533	25	1.56E-73	84.48%	5	P:starch metabolic process; F:protein binding; F:calcium ion binding; P:regulation of		GW696945

		system					flower development; C:nucleus				
Oapa1_0957_LS228_091.ab1	1	Transport protein transport Outer envelope protein 16; plastid import of protochlorophyllide oxidoreductase A; stress response	392	25	6.94E-37	75.56%	3	C:mitochondrial inner membrane; F:protein transporter activity; P:protein transport			GW696953
*Oapa1_1518_AM23_079.ab1	1	Lipid transport Lipid transfer protein precursor	597	25	9.78E-42	62.96%	2	F:lipid binding; P:lipid transport			GW697547
*arctic.fasta.screen.Contig46	3	Hormone biosynthesis or hormone induced Pollen coat protein; ABA-inducible protein-like protein	376	24	1.34E-11	66.92%	3	C:viral capsid; F:structural molecule activity; P:biological_process	-	Oapa1_0896_LS228_021.ab1; Oapa1_0539_CZ1855_048.ab1; Oapa1_0896_CZ1898_044.ab1	GW696907 GW697166 GW697294
*Oapa1_1502_AM23_	1	Dormancy/ auxin	548	25	5.03E-41	77.60%	1	P:auxin mediated signaling pathway	-		GW697540

057.ab1		repressed protein Response to biotic or abiotic factors								
arctic.fasta.screen.Contig33	2	Cold acclimation protein	828	25	1.25E-56	80.76%	1	C:membrane	-	Oapa1_1248_EY0016_045.ab1; Oapa1_0139_HFY329_008.ab1	GW697435; GW697005
Oapa1_1097_CZ1898_015.ab1	1	Defensin; PDF1	752	25	6.12E-25	87.28%	2	P:defense response to fungus; P:xenobiotic metabolic process			GW697364
arctic.fasta.screen.Contig50	6	Defensin protein; gamma-thionin/defensin	598	13	3.51E-07	71.62%	3	P:defense response; C:extracellular region; P:pathogenesis	-	Oapa1_1160_CZ1898_056.ab1; Oapa1_0496_CZ1855_001.ab1; Oapa1_0057_HFY327_035.ab1; Oapa1_1254_EY0016_037.ab1; Oapa1_0919_LS228_040.ab1; Oapa1_0919_CZ1898_058.ab1	GW697396; GW697147; GW696979; GW697437; - -
arctic.fasta.screen.Contig51	7	Defensin protein; gamma-thionin/defensin	598	14	4.58E-07	72.64%	4	P:defense response; C:extracellular region; P:defense response to bacterium; P:pathogenesis	-	Oapa1_0768_CZ1898_034.ab1; Oapa1_0108_HFY327_089.ab1; Oapa1_0929_LS228_051.ab1; Oapa1_0668_CZ1855_053.ab1; Oapa1_1137_CZ1898_033.ab1; Oapa1_0192_HFY329_062.ab1;	GW697253; GW696994; GW696928; GW697212; GW697386; GW69702

									Oapa1_0794_CZ1898_050.ab1	3 GW697260
arctic.fasta.screen.Contig26	2	Defensin protein; gamma-thionin/defensin	589	15	2.57E-07	71.80%	4	P:defense response; C:extracellular region; P:defense response to bacterium; P:pathogenesis	-	Oapa1_0945_LS228_067.ab1; Oapa1_0499_CZ1855_014.ab1	GW696944 GW697149
arctic.fasta.screen.Contig41	3	Defensin protein; gamma-thionin/defensin	611	25	1.90E-11	67.52%	2	P:response to other organism; P:defense response	-	Oapa1_1377_EY0016_048.ab1; Oapa1_0359_CZ1835_084.ab1; Oapa1_1492_AM23_044.ab1	GW697489 GW697094 -
arctic.fasta.screen.Contig24	2	Defensin; gamma-thionin/defensin; Probable protease inhibitor	350	25	8.73E-11	66.52%	1	P:defense response	-	Oapa1_0508_CZ1855_029.ab1; Oapa1_0660_CZ1855_038.ab1	GW697152 GW697209
arctic.fasta.screen.Contig38	2	Disease resistance response protein; Gamma-thionin family	926	25	1.48E-27	85.12%	2	P:defense response; P:response to biotic stimulus		Oapa1_0737_CZ1898_027.ab1; Oapa1_0657_CZ1855_044.ab1	GW697243 GW697207
arctic.fasta.screen.Contig15	1	Disease resistance response protein; Gamma-thionin family	456	25	8.83E-24	83.80%	1	P:defense response		Oapa1_0256_CZ1835_012.ab1	GW697045
arctic.fasta.	16	Disease	517	25	4.10E-	86.64%	2	P:defense response;		Oapa1_0917_CZ1898_062.ab	GW69730

screen.Contig56		resistance response protein; PDF1			28			P:response to biotic stimulus	1; Oapa1_1102_CZ1898_007.ab1; Oapa1_1032_CZ1898_048.ab1; Oapa1_0917_LS228_044.ab1; Oapa1_0381_CZ1835_004.ab1; Oapa1_0046_HFY327_022.ab1; Oapa1_1300_EY0016_072.ab1; Oapa1_0710_CZ1855_094.ab1; Oapa1_0878_LS228_009M13F.ab1; Oapa1_1269_EY0016_063.ab1; Oapa1_1381_EY0016_042.ab1; Oapa1_0878_CZ1898_043.ab1; Oapa1_1156_CZ1898_064.ab1; Oapa1_0554_CZ1855_053.ab1; Oapa1_0765_CZ1898_040.ab1; Oapa1_1001_CZ1898_004.ab1	0 GW697367 GW697331 GW696920 GW697103 GW696974 GW697460 GW697231 GW696873 GW697444 GW697490 GW697288 GW697395 GW697168 GW697252 GW697322
arctic.fasta.screen.Contig3	1	ELIP early light inducible protein	521	25	4.76E-48	70.40%	1	C:membrane	Oapa1_0727_CZ1898_014.ab1	GW697240

arctic.fasta.screen.Contig31	2	ELIP early light inducible protein	465	25	2.35E-48	73.08%	1	C:cell part	Oapa1_0806_CZ1898_078.ab1; Oapa1_0828_CZ1898_090.ab1	GW697265 GW697273
arctic.fasta.screen.Contig17	1	ELIP early light inducible protein	331	25	8.04E-25	87.32%	2	C:integral to membrane; C:chloroplast	Oapa1_1116_CZ1898_027.ab1	GW697373
arctic.fasta.screen.Contig36	2	Pathogenesis related protein; ; class 10; ABA-responsive protein ABR18	674	25	2.75E-68	86.20%	3	P:defense response; P:response to biotic stimulus; C:cytoplasm	Oapa1_0825_CZ1898_094.ab1; Oapa1_0694_CZ1855_076.ab1	GW697272 GW697224
*arctic.fasta.screen.Contig34	2	Pathogenesis related protein; ABA-responsive protein ABR18	703	25	1.35E-68	85.24%	3	P:defense response; P:response to biotic stimulus; C:cytoplasm	Oapa1_0910_LS228_041.ab1; Oapa1_0910_CZ1898_055.ab1	GW696917 GW697298
*arctic.fasta.screen.Contig39	2	Pathogenesis related protein; ABA-responsive protein ABR18	661	25	1.71E-67	85.52%	3	P:defense response; P:response to biotic stimulus; C:cytoplasm	Oapa1_0570_CZ1855_079.ab1; Oapa01_0010_CZ1298_034.ab1	GW697174 GW697563
*arctic.fasta.screen.Contig9	1	Pathogenesis related protein; ABA-responsive	649	25	1.66E-59	84.32%	3	P:defense response; P:response to biotic stimulus; C:cytoplasm	Oapa1_0060_HFY327_048.ab1	GW696980

		protein ABR18								
*arctic.fasta.screen.Contig61	94	Pathogenesis related protein; class 10	749	25	1.11E-66	85.48%	3	P:defense response; P:response to biotic stimulus; C:cytoplasm	Oapa1_0068_HFY327_063.ab1; Oapa1_0961_LS228_083.ab1; Oapa01_0003_CZ1298_017.ab1; Oapa1_0473_CZ1835_083.ab1; Oapa1_1410_EY0016_073.ab1; Oapa1_0338_CZ1835_072.ab1; Oapa1_0449_CZ1835_079.ab1; Oapa1_1356_EY0016_019.ab1; Oapa1_0025_HFY327_008.ab1; Oapa1_0962_LS228_081.ab1; Oapa1_1151_CZ1898_055.ab1; Oapa1_1198_CZ1898_090.ab1; Oapa1_1320_EY0016_092.ab1; Oapa1_1184_CZ1898_095.ab1; Oapa1_1420_EY0016_072.ab1; Oapa1_0478_CZ1835_092.ab1; Oapa1_0431_CZ1835_059.ab1; Oapa1_0364_CZ1835_013.ab1;	GW696985; GW696955; GW697560; GW697139; GW697504; GW697083; GW697130; GW697479; GW696967; GW696956; GW697392; GW697417; GW697463; GW697412; GW697510; GW697142; GW697126;

Oapa1_0466_CZ1835_066.ab1;	GW697096
Oapa1_0268_CZ1835_025.ab1;	GW697135
Oapa1_0196_HFY329_054.ab1;	GW697049
Oapa1_0225_HFY329_093.ab1;	GW697026
Oapa1_0673_CZ1855_060.ab1;	GW697035
Oapa1_0709_CZ1855_096.ab1;	GW697214
Oapa1_0240_HFY329_088.ab1;	GW697230
Oapa1_1406_EY0016_050.ab1;	GW697040
Oapa1_0067_HFY327_034.ab1;	GW697501
Oapa1_0243_HFY329_082.ab1;	GW696984
Oapa1_0711_CZ1855_092.ab1;	GW697041
Oapa1_0342_CZ1835_066.ab1;	GW697232
Oapa1_0603_CZ1855_092.ab1;	GW697085
Oapa1_0160_HFY329_020.ab1;	GW697182
Oapa1_0605_CZ1855_088.ab1;	GW697011
Oapa1_0696_CZ1855_072.ab1;	GW697183
Oapa1_0915_LS228_048.ab1;	GW697226
Oapa1_1222_EY0016_008.ab1;	GW696919
Oapa1_1466_AM23_025.ab1;	

Oapa1_0369_CZ1835_003.ab1;	GW697422
Oapa1_0512_CZ1855_023.ab1;	GW697526
Oapa1_0465_CZ1835_068.ab1;	GW697099
Oapa1_0562_CZ1855_060.ab1;	GW697154
Oapa1_0494_CZ1855_005.ab1;	GW697134
Oapa1_0348_CZ1835_087.ab1;	GW697171
Oapa1_1273_EY0016_055.ab1;	GW697146
Oapa1_0282_CZ1835_045.ab1;	GW697088
Oapa1_1298_EY0016_076.ab1;	GW697445
Oapa1_1175_CZ1898_065.ab1;	GW697058
Oapa1_0382_CZ1835_002.ab1;	GW697458
Oapa1_1236_EY0016_019.ab1;	GW697404
Oapa1_0987_CZ1898_011.ab1;	GW697104
Oapa1_1262_EY0016_044.ab1;	GW697427
Oapa1_1124_CZ1898_028.ab1;	GW697317
Oapa1_0652_CZ1855_037.ab1;	GW697439
Oapa1_1108_CZ1898_012.ab1;	GW697379
Oapa1_0492_CZ1855_009.ab1;	GW697203

Oapa1_0612_CZ1855_009.ab1;	GW697370
Oapa1_0632_CZ1855_023.ab1;	GW697144
Oapa1_0895_0.4uLdGTP_JF1427_080.ab1;	GW697185
Oapa1_0531_CZ1855_043.ab1;	GW697195
Oapa1_1098_CZ1898_013.ab1;	GW696900
Oapa1_0457_CZ1835_065.ab1;	GW697161
Oapa1_0895_JB981_056PCR_M13F.ab1;	GW697365
Oapa1_0895_betaine_1uLBDT_JF1427_030.ab1;	GW697132
Oapa1_1360_EY0016_032.ab1;	GW696905
Oapa1_0061_HFY327_046.ab1;	GW696903
Oapa1_0954_LS228_066.ab1;	GW697481
Oapa1_0895_1uLdGTP_JF1427_032.ab1;	GW696981
Oapa1_1291_EY0016_069.ab1;	GW696950
Oapa1_0870_CZ1898_022.ab1;	GW696901
Oapa1_0960_LS228_085.ab1;	GW697456
Oapa1_0685_CZ1855_073.ab1;	GW697286
Oapa1_1371_EY0016_041.ab1;	GW696954
Oapa1_0400_CZ1835_018.ab1;	GW697220
Oapa1_1444_AM23_015.ab1;	
Oapa1_1141_CZ1898_042.ab	

1; Oapa1_0250_CZ1835_007.ab	GW697486
1; Oapa1_0792_CZ1898_052.ab	GW697117
1; Oapa1_1031_CZ1898_033.ab	GW697516
1; Oapa1_1045_CZ1898_055.ab	GW697388
1; Oapa1_0796_CZ1898_077.ab	GW697042
1; Oapa1_0981_CZ1898_088.ab	GW697259
1; Oapa1_0895_CH0009_066.ab	GW697330
1; Oapa1_0837_CZ1898_007.ab	GW697337
1; Oapa1_0127_HFY329_013.ab	GW697261
1; Oapa1_0121_HFY327_090.ab	GW697315
1; Oapa1_0231_HFY329_083.ab	GW696904
1; Oapa1_0230_HFY329_085.ab	GW697276
1; Oapa1_1055_CZ1898_052.ab	GW696999
1; Oapa1_0507_CZ1855_031.ab	GW696997
1; Oapa1_0437_CZ1835_049.ab	GW697037
1; Oapa1_0804_CZ1898_065.ab	GW697036
1; Oapa1_0674_CZ1855_058.ab	GW697342 -
1; Oapa1_0356_CZ1835_090.ab	GW697128

									1; Oapa1_1120_CZ1898_019.ab1	GW69726 4 GW69721 5 GW69709 2 GW69737 6
arctic.fasta.screen.Contig2	1	Pathogenesis related protein; class 10	685	25	2.87E-68	84.48%	3	P:defense response; P:response to biotic stimulus; C:cytoplasm	Oapa1_0775_CZ1898_055.ab1	GW69725 4
arctic.fasta.screen.Contig14	1	Pathogenesis related protein; class 10	643	25	5.56E-68	85.76%	3	P:defense response; P:response to biotic stimulus; C:cytoplasm	Oapa1_1183_CZ1898_066.ab1	GW69741 1
arctic.fasta.screen.Contig13	1	Pathogenesis related protein; class 10	638	25	4.91E-69	85.52%	3	P:defense response; P:response to biotic stimulus; C:cytoplasm	Oapa1_1193_CZ1898_081.ab1	GW69741 6
arctic.fasta.screen.Contig43	3	Pathogenesis related protein; class 10	695	25	9.38E-59	85.60%	3	P:defense response; P:response to biotic stimulus; C:cytoplasm	Oapa1_1484_AM23_039.ab1; Oapa1_0567_CZ1855_052.ab1; Oapa1_1122_CZ1898_032.ab1	GW69753 6 GW69717 3 GW69737 8
*arctic.fasta.screen.Contig18	1	Pathogenesis related protein; class 10	679	25	1.39E-67	85.04%	3	P:defense response; P:response to biotic stimulus; C:cytoplasm	Oapa1_1061_CZ1898_071.ab1	GW69734 7
Oapa1_1418_EY0016_076.ab1	1	Pathogenesis related basic PR1	768	25	1.05E-59	76.64%	1	C:extracellular region		GW69750 8

arctic.fasta.screen.Contig16	1	Pathogenesis-related protein 4a	702	25	1.22E-61	83.60%	6	P:defense response to bacterium; F:chitinase activity; P:chitin catabolic process; P:cell wall catabolic process; P:defense response to fungus; F:chitin binding	EC:3.2.1.14,	Oapa1_1132_CZ1898_043.ab1	GW697383
arctic.fasta.screen.Contig10	1	Pathogenesis-related protein 4a	586	25	1.98E-60	82.92%	6	P:defense response to bacterium; F:chitinase activity; P:chitin catabolic process; P:cell wall catabolic process; P:defense response to fungus; F:chitin binding	EC:3.2.1.14,	Oapa1_0318_CZ1835_079.ab1	GW697076
Oapa1_1243_EY0016_024.ab1	1	Pathogenesis-related protein 5	692	25	7.21E-67	67.76%	2	C:cytoplasmic membrane-bounded vesicle; P:response to stimulus	-		GW697432
arctic.fasta.screen.Contig35	2	Protease inhibitor	479	25	1.19E-28	78.08%	3	F:peptidase activity; F:serine-type endopeptidase inhibitor activity; P:response to wounding	-	Oapa1_0825_CZ1898_094.ab1; Oapa1_0694_CZ1855_076.ab1	GW697272 GW697224
arctic.fasta.screen.Contig12	1	Cold-regulated or drought induced protein; KS-type dehydrin	386	16	4.72E-09	59.38%	8	P:response to water; P:response to stress; F:neurotransmitter:sodium symporter activity; C:membrane; P:neurotransmitter transport; C:integral	-	Oapa1_0267_CZ1835_027.ab1	GW697048

								to plasma membrane; F:transporter activity; P:transport			
arctic.fasta.screen.Contig27	2	SRC1 or cold-acclimation specific protein 15 BudCar15	455	3	1.80E-09	61.33%	2	P:response to water; P:response to stress	-	Oapa1_0752_CZ1898_047.ab1; Oapa1_0429_CZ1835_063.ab1	GW697248 GW697124
arctic.fasta.screen.Contig59	45	Cold-regulated or drought induced protein	744	0			0		-	Oapa1_0719_CZ1898_013.ab1; Oapa1_1100_CZ1898_009.ab1; Oapa1_1375_EY0016_035.ab1; Oapa1_0584_CZ1855_072.ab1; Oapa1_0475_CZ1835_096.ab1; Oapa1_1352_EY0016_021.ab1; Oapa1_0178_HFY329_063.ab1; Oapa1_0407_CZ1835_043.ab1; Oapa1_1373_EY0016_039.ab1; Oapa1_1136_CZ1898_035.ab1; Oapa1_1396_EY0016_051.ab1; Oapa1_0220_HFY329_072.ab1; Oapa1_0326_CZ1835_067.ab	GW697235 GW697366 GW697488 GW697177 GW697140 GW697478 GW697017 GW697119 - GW697385 GW697497 GW697034 GW697078 GW69745

1;	3
Oapa1_1283_EY0016_052.ab	GW69716
1;	7
Oapa1_0552_CZ1855_055.ab	GW69744
1;	3
Oapa1_1268_EY0016_034.ab	GW69707
1;	9
Oapa1_0327_CZ1835_065.ab	GW69701
1;	8
Oapa1_0180_HFY329_059.ab	GW69702
1;	0
Oapa1_0184_HFY329_053.ab	GW69733
1;	5
Oapa1_1042_CZ1898_061.ab	GW69706
1;	7
Oapa1_0299_CZ1835_034.ab	GW69707
1;	1
Oapa1_0306_CZ1835_055.ab	GW69755
1;	7
Oapa1_1543_AM23_085.ab1;	GW69718
Oapa1_0614_CZ1855_005.ab	7
1;	GW69735
Oapa1_1071_CZ1898_074.ab	2
1;	GW69689
Oapa1_0891_LS228_031.ab1;	6
Oapa1_1253_EY0016_039.ab	GW69743
1;	6
Oapa1_0683_CZ1855_077.ab	GW69721
1;	9
Oapa1_0695_CZ1855_074.ab	GW69722
1;	5
Oapa1_0952_LS228_070.ab1;	GW69694
Oapa1_0693_CZ1855_078.ab	8
1;	GW69722
Oapa1_0514_CZ1855_019.ab	3
1;	GW69715

									Oapa1_1215_EY0016_001.ab1; Oapa1_0153_HFY329_017.ab1; Oapa1_0395_CZ1835_028.ab1; Oapa1_0951_LS228_072.ab1; Oapa1_1279_EY0016_060.ab1; Oapa1_0812_CZ1898_068.ab1; Oapa1_0390_CZ1835_017.ab1; Oapa1_0907_LS228_047.ab1; Oapa1_0928_LS228_053.ab1; Oapa1_1025_CZ1898_043.ab1; Oapa1_0036_HFY327_021.ab1; Oapa1_0346_CZ1835_089.ab1; Oapa1_1421_EY0016_070.ab1	5 GW697419 GW697009 GW697112 GW696947 GW697450 GW697267 GW697109 GW696915 GW696927 GW697329 GW696970 GW697087 GW697511
arctic.fasta.screen.Contig25	2	Cold-acclimation specific protein 15 BudCar15	703	1	1.81E-06	67.00%	0	-	Oapa1_0942_LS228_073.ab1; Oapa1_0943_LS228_071.ab1	GW696941 GW696942
Oapa1_0752_CZ1898_047.ab1	1	Cold-regulated or drought induced	428	0			0	-		GW697248

		protein									
arctic.fasta.screen.Contig44	3	Cold-regulated or drought induced protein;	551	9	3.89E-11	67.56%	2	P:response to water; P:response to stress	-	Oapa1_0717_CZ1855_082.ab1; Oapa1_0937_LS228_052.ab1; Oapa1_1460_AM23_006.ab1	GW697234 GW696934 GW697523
Oapa1_0429_CZ1835_063.ab1	1	Cold-regulated or drought induced protein; KS-type dehydrin	447	6	1.13E-08	66.00%	2	P:response to water; P:response to stress	-		GW697124
arctic.fasta.screen.Contig7	1	Cold-regulated or drought induced protein; Bud cold acclimation responsive protein 5	578	12	1.07E-10	52.08%	2	P:response to water; P:response to stress	-	Oapa1_0468_CZ1835_093.ab1	GW697136
arctic.fasta.screen.Contig23	1	Cold-regulated or drought induced protein; cold-acclimation specific protein 15	695	5	6.93E-09	50.60%	2	P:response to water; P:response to stress	-	Oapa1_1278_EY0016_062.ab1	GW697449

arctic.fasta.screen.Contig21	1	Cold-regulated or drought induced protein; cold-induced protein CIP	554	25	3.29E-19	53.92%	9	C:membrane; F:protein serine/threonine kinase activity; P:response to stress; P:protein homooligomerization; F:protein kinase activity; P:protein amino acid phosphorylation; F:ATP binding; P:response to water; F:kinase activity	-	Oapa1_0930_LS228_049.ab1	GW696929
arctic.fasta.screen.Contig47	4	Cold-regulated or drought induced protein; dehydrin family protein	800	0			0		-	Oapa1_0883_JF1425_008M13R.ab1; Oapa1_1467_AM23_023.ab1; Oapa1_1470_AM23_017.ab1; Oapa1_0883_LS228_016M13F.ab1	GW696883 GW697527 GW697529 GW696884
Oapa1_0403_CZ1835_047.ab1	1	Drought-stressed	381	2	2.96E-19	86.50%	1	C:mitochondrion			GW697118
arctic.fasta.screen.Contig5	1	Dehydrin, very weak similarity to RAB18 and ERD10 COR proteins	314	0			0		-	Oapa1_0572_CZ1855_075.ab1	GW697175
Oapa1_1329_EY0016_011.ab1	1	Drought-stress expressed	478	3	8.87E-08	74.00%	0		-		GW697468

Oapa1_0158_HFY329_024.ab1	1	Drought-stress expressed, very weak similarity to ERD10 and ERD14 COR protein	394	0			0	-		GW697010
arctic.fasta.screen.Contig54	8	LEA late embryogenesis abundant protein	659	3	2.73E-09	50.33%	0	-	Oapa1_1321_EY0016_090.ab1; Oapa1_1482_AM23_043.ab1; Oapa1_1453_AM23_001.ab1; Oapa1_1237_EY0016_017.ab1; Oapa1_0331_CZ1835_076.ab1; Oapa1_0283_CZ1835_043.ab1; Oapa1_0288_CZ1835_033.ab1; Oapa1_0471_CZ1835_087.ab1	GW697464 GW697535 GW697520 GW697428 GW697082 GW697059 GW697063 GW697138
arctic.fasta.screen.Contig1	1	LEA late embryogenesis abundant protein type 3; maturation protein PM32	557	2	1.52E-08	57.00%	0	-	Oapa1_0894_LS228_025.ab1	GW696899
arctic.fasta.	1	LEA late	537	3	6.93E-	49.00%	0	-	Oapa1_1238_EY0016_032.ab	GW69742

screen.Contig20		embryogenesis abundant protein			08					1		9
Oapa1_1119_CZ1898_021.ab1	1	LEA late embryogenesis abundant protein	671	25	4.67E-44	48.76%	2	C:plastid; P:response to stress	-			GW697375
Oapa1_0350_CZ1835_085.ab1	1	LEA late embryogenesis abundant protein; seed maturation protein PM35; COR8.5 LEA14 late embryogenesis abundant protein; seed maturation protein PM22 LEA2 late embryogenesis abundant protein; protein; seed maturation	371	14	1.59E-28	72.36%	2	F:molecular_function; P:biological_process	-			GW697089
Oapa1_1139_CZ1898_046.ab1	1		754	25	2.40E-61	78.92%	1	P:response to desiccation				GW697387
Oapa1_1165_CZ1898_050.ab1	1		626	25	1.16E-27	54.32%	2	F:molecular_function; P:embryonic development ending in seed dormancy	-			GW697399

		protein PM30									
arctic.fasta.screen.Contig28	2	Little protein 1; unknown function	501	25	6.20E-31	81.40%	0		-	Oapa1_0071_HFY327_059.ab1; Oapa1_0345_CZ1835_091.ab1	GW696986; GW697086
arctic.fasta.screen.Contig30	2	Proline-rich protein	474	25	2.58E-23	49.80%	0		-	Oapa1_0294_CZ1835_040.ab1; Oapa1_0470_CZ1835_089.ab1	GW697065; GW697137
Oapa1_0634_CZ1855_019.ab1	1	Proline-rich protein	302	0			0		-		GW697196
Oapa1_0852_CZ1898_004.ab1	1	REF rubber elongation factor family protein; stress related; endomembrane system	257	25	1.95E-18	68.28%	2	P:response to stress; P:biological_process	-		GW697279
Oapa1_0287_CZ1835_035.ab1	1	Subtilisin inhibitor; protease inhibitor family	472	17	1.25E-17	67.06%	2	F:serine-type endopeptidase inhibitor activity; P:response to wounding			GW697062
*Oapa01_0009_CZ1298_033.ab1	1	Trypsin protein inhibitor 3; Kunitz trypsin protease inhibitor Lipids	730	25	3.04E-82	54.68%	2	F:peptidase activity; F:endopeptidase inhibitor activity			GW697562

Oapa1_1044_CZ1898_057.ab1	1	Dienoyl-CoA isomerase; enoyl-CoA hydratase; propanoate, fatty acid and amino acid metabolism	310	22	3.14E-13	74.09%	2	F:catalytic activity; P:metabolic process	EC:4.2.1.17,		GW697336
Oapa1_0291_CZ1835_046.ab1	1	Pyruvate - fatty acids Acetyl-carboxylase beta subunit	290	25	1.59E-20	85.20%	0		-		GW697064
*arctic.fasta.screen.Contig58	32	Acetyl-carboxylase carboxyltransferase beta subunit	776	25	1.05E-43	91.28%	7	C:acetyl-CoA carboxylase complex; P:fatty acid biosynthetic process; F:transferase activity; C:chloroplast; F:zinc ion binding; F:acetyl-CoA carboxylase activity; P:pyruvate metabolic process	-	Oapa1_0639_CZ1855_030.ab1; Oapa1_1289_EY0016_073.ab1; Oapa1_1309_EY0016_091.ab1; Oapa1_0955_LS228_095.ab1; Oapa01_0002_CZ1298_002.ab1; Oapa1_0955_CZ1898_074.ab1; Oapa1_0210_HFY329_071.ab1; Oapa1_1362_EY0016_028.ab1; Oapa1_1519_AM23_077.ab1; Oapa1_0043_HFY327_028.ab1; Oapa1_0387_CZ1835_023.ab1;	GW697198 GW697455 GW697461 GW696951 GW697559 GW697308 GW697031 GW697482 GW697548 GW696972

Oapa1_0235_HFY329_092.ab1;	GW69710
Oapa1_0281_CZ1835_047.ab1;	GW69703
Oapa1_0680_CZ1855_079.ab1;	GW69705
Oapa1_1417_EY0016_078.ab1;	GW69721
Oapa1_0415_CZ1835_035.ab1;	GW69750
Oapa1_0148_HFY329_027.ab1;	GW69712
Oapa1_0174_HFY329_040.ab1;	GW69700
Oapa1_1259_EY0016_048.ab1;	GW69701
Oapa1_0908_LS228_045.ab1;	GW69743
Oapa1_0656_CZ1855_046.ab1;	GW69691
Oapa1_0815_CZ1898_093.ab1;	GW69720
Oapa1_0746_CZ1898_028.ab1;	GW69726
Oapa1_0854_CZ1898_031.ab1;	GW69724
Oapa1_1005_CZ1898_027.ab1;	GW69728
Oapa1_0908_CZ1898_057.ab1;	GW69732
Oapa1_1019_CZ1898_024.ab1;	GW69729
Oapa1_1052_CZ1898_058.ab1;	GW69732
Oapa1_0564_CZ1855_056.ab1;	GW69734
Oapa1_1400_EY0016_062.ab	0

*arctic.fasta.screen.Contig57	30	Acetyl-carboxylase carboxyltransferase beta subunit	733	25	3.68E-53	90.76%	7	C:acetyl-CoA carboxylase complex; P:fatty acid biosynthetic process; F:transferase activity; C:chloroplast; F:zinc ion binding; F:acetyl-CoA carboxylase activity; P:pyruvate metabolic process	EC:6.4.1.2,	1; Oapa1_1277_EY0016_064.ab1; Oapa1_1347_EY0016_027.ab1 GW69717 2 GW69749 8 GW69744 8 GW69747 6 Oapa1_1229_EY0016_027.ab1; Oapa1_0758_CZ1898_037.ab1; Oapa1_1394_EY0016_055.ab1; Oapa1_0328_CZ1835_080.ab1; Oapa1_1338_EY0016_010.ab1; Oapa1_0534_CZ1855_037.ab1; Oapa1_1429_EY0016_087.ab1; Oapa1_0362_CZ1835_015.ab1; Oapa1_1532_AM23_074.ab1; Oapa01_0001_CZ1298_001.ab1; Oapa1_0532_CZ1855_041.ab1; Oapa1_1217_EY0016_014.ab1; Oapa1_1242_EY0016_026.ab1; Oapa1_0198_HFY329_052.ab1; Oapa1_0720_CZ1898_011.ab GW69742 6 GW69725 0 GW69749 6 GW69708 0 GW69747 2 GW69716 3 GW69751 2 GW69709 5 GW69755 2 GW69755 8 GW69716 2 GW69742 1 GW69743 1 GW69702 7

Ribosome and translation	1; Oapa1_0644_CZ1855_020.ab1; Oapa1_0262_CZ1835_002.ab1; Oapa1_0137_HFY329_012.ab1; Oapa1_1480_AM23_047.ab1; Oapa1_0790_CZ1898_054.ab1; Oapa1_1346_EY0016_029.ab1; Oapa1_0734_CZ1898_031.ab1; Oapa1_1034_CZ1898_044.ab1; Oapa1_0830_CZ1898_086.ab1; Oapa1_0168_HFY329_035.ab1; Oapa1_1452_AM23_003.ab1; Oapa1_1475_AM23_024.ab1; Oapa1_0978_CZ1898_092.ab1; Oapa1_1517_AM23_050.ab1; Oapa1_1401_EY0016_060.ab1	GW697236 GW697201 GW697047 GW697003 GW697534 GW697258 GW697475 GW697242 GW697332 GW697274 GW697015 GW697519 GW697532 GW697314 GW697546 GW697499

		Ribosome biogenesis and assembly - organelles									
Oapa1_1322_EY0016_088.ab1	1	30S ribosomal protein	696	25	4.64E-45	69.80%	5	F:structural constituent of ribosome; P:translation; C:plastid; C:ribosome; P:ribosome biogenesis and assembly	EC:3.6.5.3,		GW697465
*arctic.fasta.screen.Contig6	1	30S ribosomal protein S12 B	473	25	2.95E-43	97.68%	7	F:structural constituent of ribosome; C:small ribosomal subunit; C:chloroplast; P:translation; F:rRNA binding; P:response to antibiotic; P:ribosome biogenesis and assembly	EC:3.6.5.3,	Oapa1_0504_CZ1855_006.ab1	GW697151
*arctic.fasta.screen.Contig32	2	30S ribosomal protein s12	420	25	1.52E-63	97.64%	7	F:structural constituent of ribosome; C:small ribosomal subunit; C:chloroplast; P:translation; F:rRNA binding; P:response to antibiotic; P:ribosome	EC:3.6.5.3,	Oapa1_0747_CZ1898_026.ab1; Oapa1_1445_AM23_013.ab1	-GW697517

								biogenesis and assembly			
Oapa1_1456_AM23_014.ab1	1	30S ribosomal protein S18; rps18	429	25	1.96E-35	90.80%	6	C:ribosome; F:structural constituent of ribosome; C:chloroplast; P:translation; F:rRNA binding; P:ribosome biogenesis and assembly	EC:3.6.5.3,		GW697521
*arctic.fasta.screen.Contig49	5	23S ribosomal RNA; protein orf91; chloroplast genome	616	25	7.12E-38	74.92%	1	C:chloroplast		Oapa1_0890_0.4uLdGTP_JF1427_096.ab1; Oapa1_0890_1uLdGTP_JF1427_048.ab1; Oapa1_0890_LS228_002.ab1; Oapa1_0890_betaine_1uLBDT_JF1427_046.ab1; Oapa1_0890_betaine_0.5uLBDT_JF1427_094.ab1	GW696891 GW696892 GW696895 GW696894 GW696893
Oapa1_1339_EY0016_008.ab1	1	ribosomal protein L9	426	10	1.34E-19	81.30%	5	F:structural constituent of ribosome; P:ribosome biogenesis and assembly; P:translation; C:mitochondrion; C:ribosome	EC:3.6.5.3,		GW697473
		ribosome biogenesis and									

Oapa1_1144_CZ1898_036.ab1	1	assembly - cell 40S ribosomal protein similar to S12 (RPS12C); L7Ae/L30e/S12e/Gadd45 family	700	25	5.34E-57	89.84%	4	C:ribosome; F:structural constituent of ribosome; P:translation; P:ribosome biogenesis and assembly	-	GW697390
Oapa1_1038_CZ1898_038.ab1	1	40S ribosomal protein s30	398	25	6.46E-22	92.72%	5	P:protein modification process; F:structural constituent of ribosome; C:cytosolic small ribosomal subunit; P:translation; P:ribosome biogenesis and assembly	EC:3.6.5.3,	GW697334
Oapa1_0051_HFY327_047.ab1	1	60S ribosomal protein L8 (RPL8C)	669	25	1.57E-116	95.24%	7	C:nucleolus; C:vacuole; F:structural constituent of ribosome; C:plasma membrane; C:cytosolic large ribosomal subunit; P:translation; P:ribosome biogenesis and assembly	EC:3.6.5.3,	GW696977
		RNA associated								

Oapa1_0732_CZ1898_004.ab1	1	proteins EIF-4 middle initiation factor; nucleus, RNA cap binding mediating U snRNA export from the nucleus; response to abscissic acid	440	10	2.32E-16	84.00%	0		-		GW697241
arctic.fasta.screen.Contig8	1	RNA recognition motif (RRM)-containing protein	650	25	6.80E-29	76.44%	3	F:nucleic acid binding; F:nucleotide binding; F:oxidoreductase activity		Oapa1_0339_CZ1835_070.ab1	GW697084
Oapa1_1485_AM23_037.ab1	1	RNA-dependent RNA polymerase; similar to virus genomes	603	11	2.37E-75	63.73%	1	F:RNA-directed RNA polymerase activity	-		GW697537
*Oapa1_0163_HFY329_045.ab1	1	Histones Weakely similar to a Histone deacetylase and COR8.6	576	0			0		-		GW697013

		protein									
		ROS scavenging and signalling									
Oapa1_0451_CZ1835_075.ab1	1	Glutathione S-transferase	619	25	2.49E-83	80.32%	4	P:response to stress; F:glutathione transferase activity; P:glutathione metabolic process; P:glutathione conjugation reaction	EC:2.5.1.18,		GW697131
*arctic.fasta.screen.Contig29	2	Phospholipid hydroperoxide glutathione peroxidase	714	25	3.11E-84	90.16%	7	F:phospholipid-hydroperoxide glutathione peroxidase activity; F:glutathione peroxidase activity; C:cytoplasm; P:oxidation reduction; P:response to oxidative stress; P:glutathione metabolic process; P:peroxidase reaction	EC:1.11.1.12, EC:1.11.1.9,	Oapa1_0892_LS228_029.ab1; Oapa1_1131_CZ1898_045.ab1	GW696897 GW697382
Oapa1_0284_CZ1835_041.ab1	1	Lactoylglutathione lyase; glyoxalase I family protein; pyruvate metabolism	386	24	3.47E-36	67.83%	1	P:metabolic process			GW697060

		; endomembrane system								
*Oapa1_0354_CZ1835_094.ab1	1	Rhicadhesin receptor precursor (Germin-like protein); Cupin domain	681	25	5.18E-70	80.04%	11	F:nutrient reservoir activity; F:receptor activity; F:manganese ion binding; C:cytoplasmic membrane-bounded vesicle; C:cell wall; F:superoxide dismutase activity; P:oxidation reduction; C:nucleus; C:apoplast; P:signal transduction; P:superoxide metabolic process		GW697091
Oapa1_1393_EY0016_057.ab1	1	Unclassified Embryo-specific 3; located in membranes	761	25	2.13E-65	74.16%	1	C:anchored to membrane	-	GW697495
Oapa1_0925_LS228_059.ab1	1	Haloacid dehalogenase-like hydrolase family protein	710	25	1.29E-50	77.80%	2	P:metabolic process; F:hydrolase activity		GW696925
Oapa1_1349_EY0016_025.ab1	1	Chloroplast genome	273	0			0		-	GW697477

		Unknown						
arctic.fasta.screen.Contig63	263	NA	569	0	-	0	-	Oapa1_0976_CZ1898_096.ab1; Oapa1_1364_EY0016_024.ab1; Oapa1_1336_EY0016_014.ab1; Oapa1_0138_HFY329_010.ab1; Oapa1_0847_CZ1898_012.ab1; Oapa1_1530_AM23_078.ab1; Oapa1_0522_CZ1855_024.ab1; Oapa1_0518_CZ1855_030.ab1; Oapa1_1134_CZ1898_039.ab1; Oapa1_0486_CZ1855_015.ab1; Oapa1_0619_CZ1855_012.ab1; Oapa1_0447_CZ1835_052.ab1; Oapa1_1419_EY0016_074.ab1; Oapa1_1020_CZ1898_022.ab1; Oapa1_0560_CZ1855_064.ab1; Oapa1_0903_LS228_024.ab1; Oapa1_0636_CZ1855_017.ab1; Oapa1_1389_EY0016_063.ab1; Oapa1_0476_CZ1835_094.ab	GW697313; GW697483; GW697470; GW697004; GW697277; GW697551; GW697158; GW697157; GW697384; GW697143; GW697190; GW697129; GW697509; GW697327; GW697170; GW696913; GW697197; GW69749

1;	3
Oapa1_0517_CZ1855_032.ab	GW69714
1;	1
Oapa1_0968_SB1906_060.ab1	GW69715
;	6
Oapa1_1523_AM23_071.ab1;	GW69696
Oapa1_0629_CZ1855_027.ab	2
1;	GW69755
Oapa1_0692_CZ1855_080.ab	0
1;	GW69719
Oapa1_0199_HFY329_050.ab	3
1;	GW69722
Oapa1_0650_CZ1855_041.ab	2
1;	GW69702
Oapa1_1402_EY0016_058.ab	8
1;	GW69720
Oapa1_0871_CZ1898_020.ab	2 -
1;	GW69728
Oapa1_0670_CZ1855_049.ab	7
1;	GW69721
Oapa1_1057_CZ1898_079.ab	3
1;	GW69734
Oapa1_0654_CZ1855_033.ab	3
1;	GW69720
Oapa1_1479_AM23_018.ab1;	5
Oapa1_1515_AM23_052.ab1;	GW69753
Oapa1_1308_EY0016_093.ab	3
1;	GW69754
Oapa1_0659_CZ1855_040.ab	5 -
1;	GW69720
Oapa1_0893_LS228_027.ab1;	8
Oapa1_0593_CZ1855_091.ab	GW69689
1;	8
Oapa1_1146_CZ1898_063.ab	GW69718
1;	0
Oapa1_0615_CZ1855_003.ab	GW69739

1; Oapa1_0626_CZ1855_002.ab1;	1 GW697188
Oapa1_0939_CZ1898_069.ab1;	GW697192
Oapa1_0581_CZ1855_078.ab1;	GW697304
Oapa1_0497_CZ1855_016.ab1;	GW697176
Oapa1_0821_CZ1898_085.ab1;	GW697148
Oapa1_0501_CZ1855_010.ab1;	GW697271
Oapa1_0630_CZ1855_025.ab1;	GW697150
Oapa1_0939_LS228_079.ab1;	GW697194
Oapa1_1112_CZ1898_004.ab1;	GW696935
Oapa1_1223_EY0016_006.ab1;	GW697371
Oapa1_0811_CZ1898_070.ab1;	GW697423
Oapa1_0941_CZ1898_065.ab1;	GW697266
Oapa1_0949_CZ1898_076.ab1;	GW697306
Oapa1_0653_CZ1855_035.ab1;	GW697307
Oapa1_0301_CZ1835_063.ab1;	GW697204
Oapa1_0949_LS228_076.ab1;	GW697068
Oapa1_1305_EY0016_068.ab1;	GW696946 -
Oapa1_0529_CZ1855_047.ab1; Oapa1_0182_HFY329_055.ab	GW69716

1; Oapa1_0941_LS228_075.ab1; Oapa1_1013_CZ1898_017.ab 1; Oapa1_0903_CZ1898_034.ab 1; Oapa1_0081_HFY327_058.ab 1; Oapa1_0849_CZ1898_008.ab 1; Oapa1_0866_CZ1898_030.ab 1; Oapa1_0708_CZ1855_081.ab 1; Oapa1_0618_CZ1855_014.ab 1; Oapa1_0029_HFY327_031.ab 1; Oapa1_0065_HFY327_038.ab 1; Oapa1_0050_HFY327_018.ab 1; Oapa1_1337_EY0016_012.ab 1; Oapa1_1188_CZ1898_089.ab 1; Oapa1_0206_HFY329_077.ab 1; Oapa1_0136_HFY329_014.ab 1; Oapa1_0190_HFY329_064.ab 1; Oapa1_0021_HFY327_014.ab 1; Oapa1_1448_AM23_007.ab1; Oapa1_1106_CZ1898_016.ab	0 GW69701 9 GW69694 0 GW69732 4 GW69729 6 GW69698 8 GW69727 8 GW69728 3 GW69722 9 GW69718 9 GW69696 9 GW69698 3 GW69697 6 GW69747 1 GW69741 4 GW69702 9 GW69700 2 GW69702 2 GW69696

1;	5
Oapa1_1539_AM23_093.ab1;	GW69751
Oapa1_1064_CZ1898_065.ab	8
1;	GW69736
Oapa1_1323_EY0016_086.ab	8
1;	GW69755
Oapa1_1077_CZ1898_095.ab	5
1;	GW69734
Oapa1_0867_CZ1898_028.ab	9
1;	GW69746
Oapa1_1266_EY0016_038.ab	6
1;	GW69735
Oapa1_1392_EY0016_059.ab	7
1;	GW69728
Oapa1_0608_CZ1855_082.ab	4
1;	GW69744
Oapa1_1024_CZ1898_045.ab	2
1;	GW69749
Oapa1_1015_CZ1898_030.ab	4
1;	GW69718
Oapa1_0166_HFY329_039.ab	4
1;	GW69732
Oapa1_1403_EY0016_056.ab	8
1;	GW69732
Oapa1_0973_CZ1898_085.ab	5
1;	GW69701
Oapa1_1244_EY0016_022.ab	4
1;	GW69750
Oapa1_0059_HFY327_033.ab	0
1;	GW69731
Oapa1_0090_HFY327_071.ab	2
1;	GW69743
Oapa1_0179_HFY329_061.ab	3 - - -
1;	GW69729
Oapa1_0885_CZ1898_033.ab	2
1;	GW69693

Oapa1_0940_betaine_0.5uLBDT_JF1427_077.ab1;	8 - GW69688
Oapa1_0026_HFY327_006.ab1;	6 - - GW69748
Oapa1_0884_1uLdGTP_JF1427_016.ab1;	0 - - - GW69693
Oapa1_0222_HFY329_068.ab1;	9 GW69720
Oapa1_0079_HFY327_060.ab1;	0 - GW69688
Oapa1_1359_EY0016_017.ab1;	7 GW69729
Oapa1_0066_HFY327_036.ab1;	9 - GW69711
Oapa1_0075_HFY327_051.ab1;	3 GW69731
Oapa1_0228_HFY329_089.ab1;	1 GW69722
Oapa1_0940_betaine_1uLBDT_JF1427_029.ab1;	1 GW69712
Oapa1_0643_CZ1855_022.ab1;	3 GW69691
Oapa1_0916_LS228_046.ab1;	2 GW69688
Oapa1_0884_betaine_1uLBDT_JF1427_014.ab1;	5 GW69710
Oapa1_0916_CZ1898_064.ab1;	2 GW69693
Oapa1_0149_HFY329_025.ab1;	1 - - GW69692
Oapa1_0396_CZ1835_026.ab1;	6 - GW69734
Oapa1_0967_CZ1898_089.ab1;	4 GW69721
Oapa1_0687_CZ1855_069.ab1;	1 GW69754
Oapa1_0425_CZ1835_036.ab	

1;	4
Oapa1_0899_betaine_1uLBDT_JF1427_013.ab1;	GW69734
	1
Oapa1_0884_0.4uLdGTP_JF1427_064.ab1;	GW69738
	0
Oapa1_0379_CZ1835_008.ab1;	GW69740
	3
Oapa1_0933_LS228_060.ab1;	GW69693
Oapa1_0884_LS228_014M13F.ab1;	0
	GW69753
Oapa1_0545_CZ1855_038.ab1;	8
	GW69729
Oapa1_0927_LS228_055.ab1;	0
Oapa1_1037_CZ1898_040.ab1;	GW69732
	1
Oapa1_1058_CZ1898_077.ab1;	GW69721
	0
Oapa1_0666_CZ1855_057.ab1;	GW69693
	7
Oapa1_1514_AM23_054.ab1;	GW69728
Oapa1_1053_CZ1898_056.ab1;	9
	GW69733
Oapa1_1127_CZ1898_022.ab1;	3
	GW69691
Oapa1_1174_CZ1898_067.ab1;	1
	GW69730
Oapa1_0932_LS228_062.ab1;	1
Oapa1_1493_AM23_042.ab1;	GW69702
Oapa1_0881_CZ1898_039.ab1;	5
	GW69687
Oapa1_0998_CZ1898_010.ab1;	2
	GW69752
Oapa1_0662_CZ1855_034.ab1;	8
	GW69737
Oapa1_0940_1uLdGTP_JF1427_031.ab1;	2
	GW69739

Oapa1_0879_CZ1898_041.ab1;	4 GW69707
Oapa1_1036_CZ1898_042.ab1;	4 GW69711
Oapa1_0899_betaine_0.5uLBDT_JF1427_061.ab1;	5 GW69706
Oapa1_0920_CZ1898_056.ab1;	1 GW69741
Oapa1_0195_HFY329_056.ab1;	0 GW69754
Oapa1_0876_LS228_013M13F.ab1;	9 GW69745
Oapa1_1468_AM23_021.ab1;	4 -
Oapa1_1114_CZ1898_031.ab1;	GW69741 5
Oapa1_1154_CZ1898_051.ab1;	GW69734 8
Oapa1_0315_CZ1835_054.ab1;	GW69744 7
Oapa1_0398_CZ1835_022.ab1;	GW69740 7 - - - - - -
Oapa1_0286_CZ1835_037.ab1;	- - - - - - - - - - - - - - - -
Oapa1_1182_CZ1898_068.ab1;	- - - - - - - - - - - - - - - -
Oapa1_1521_AM23_073.ab1;	- - - - - -
Oapa1_1284_EY0016_050.ab1;	GW69736 0 - - -
Oapa1_1529_AM23_080.ab1;	GW69723
Oapa1_1192_CZ1898_083.ab1;	3 - GW69727
Oapa1_1063_CZ1898_067.ab1;	0 GW69724
Oapa1_1276_EY0016_049.ab1;	6 GW69711
Oapa1_1179_CZ1898_074.ab	0 - - - -

1; Oapa1_1014_CZ1898_032.ab 1; Oapa1_1498_AM23_034.ab1; Oapa1_0859_CZ1898_025.ab 1; Oapa1_0884_B981_040PCR_ M13F.ab1; Oapa1_0848_CZ1898_010.ab 1; Oapa1_1528_AM23_065.ab1; Oapa1_0640_CZ1855_028.ab 1; Oapa1_0480_CZ1835_088.ab 1; Oapa1_0884_betaine_0.5uLB DT_JF1427_062.ab1; Oapa1_0577_CZ1855_067.ab 1; Oapa1_1425_EY0016_093.ab 1; Oapa1_1397_EY0016_049.ab 1; Oapa1_1423_EY0016_066.ab 1; Oapa1_0376_CZ1835_014.ab 1; Oapa1_0161_HFY329_018.ab 1; Oapa1_0541_CZ1855_046.ab 1; Oapa1_0044_HFY327_026.ab 1; Oapa1_1331_EY0016_007.ab 1; Oapa1_0011_HFY327_015.ab	GW69688 8 - - - GW69750 6 -

1;
Oapa1_0439_CZ1835_062.ab
1;
Oapa1_0707_CZ1855_083.ab
1;
Oapa1_1147_CZ1898_061.ab
1;
Oapa1_1384_EY0016_038.ab
1;
Oapa1_1413_EY0016_067.ab
1;
Oapa1_0506_CZ1855_002.ab
1;
Oapa1_0616_CZ1855_001.ab
1;
Oapa1_0623_CZ1855_006.ab
1;
Oapa1_1428_EY0016_089.ab
1;
Oapa1_1431_EY0016_083.ab
1;
Oapa1_1027_CZ1898_039.ab
1;
Oapa1_0583_CZ1855_074.ab
1;
Oapa1_0568_CZ1855_050.ab
1;
Oapa1_0366_CZ1835_009.ab
1;
Oapa1_1509_AM23_060.ab1;
Oapa1_0741_CZ1898_019.ab
1;
Oapa1_0533_CZ1855_039.ab
1;
Oapa1_0028_HFY327_002.ab
1;

Oapa1_1478_AM23_020.ab1;
Oapa1_0803_CZ1898_067.ab
1;
Oapa1_0550_CZ1855_059.ab
1;
Oapa1_1084_CZ1898_083.ab
1;
Oapa1_0012_HFY327_013.ab
1;
Oapa1_1049_CZ1898_064.ab
1;
Oapa1_1404_EY0016_054.ab
1;
Oapa1_0713_CZ1855_088.ab
1;
Oapa1_0314_CZ1835_056.ab
1;
Oapa1_0818_CZ1898_089.ab
1;
Oapa1_0749_CZ1898_022.ab
1;
Oapa1_0391_CZ1835_032.ab
1;
Oapa1_0417_CZ1835_048.ab
1;
Oapa1_0899_LS228_032.ab1;
Oapa1_0086_HFY327_079.ab
1;
Oapa1_0912_CZ1898_051.ab
1;
Oapa1_0884_JF1425_024M13
R.ab1;
Oapa1_0761_CZ1898_048.ab
1;
Oapa1_0757_CZ1898_039.ab
1;

Oapa1_1096_CZ1898_082.ab
1;
Oapa1_1415_EY0016_080.ab
1;
Oapa1_0124_HFY327_084.ab
1;
Oapa1_0123_HFY327_086.ab
1;
Oapa1_0936_CZ1898_073.ab
1;
Oapa1_1281_EY0016_056.ab
1;
Oapa1_0293_CZ1835_042.ab
1;
Oapa1_0321_CZ1835_075.ab
1;
Oapa1_0786_CZ1898_060.ab
1;
Oapa1_0918_CZ1898_060.ab
1;
Oapa1_0218_HFY329_076.ab
1;
Oapa1_0667_CZ1855_055.ab
1;
Oapa1_1028_CZ1898_037.ab
1;
Oapa1_0085_HFY327_050.ab
1;
Oapa1_0020_HFY327_016.ab
1;
Oapa1_0213_HFY329_067.ab
1;
Oapa1_0712_CZ1855_090.ab
1;
Oapa1_0072_HFY327_057.ab
1;

Oapa1_1390_EY0016_061.ab
1;
Oapa1_0850_CZ1898_006.ab
1;
Oapa1_1203_EY0016_015.ab
1;
Oapa1_1062_CZ1898_069.ab
1;
Oapa1_0140_HFY329_006.ab
1;
Oapa1_1043_CZ1898_059.ab
1;
Oapa1_0795_CZ1898_079.ab
1;
Oapa1_0141_HFY329_004.ab
1;
Oapa1_0862_CZ1898_019.ab
1;
Oapa1_0104_HFY327_093.ab
1;
Oapa1_1017_CZ1898_028.ab
1;
Oapa1_0948_CZ1898_078.ab
1;
Oapa1_0700_CZ1855_095.ab
1;
Oapa1_0991_CZ1898_003.ab
1;
Oapa1_0367_CZ1835_007.ab
1;
Oapa1_0432_CZ1835_057.ab
1;
Oapa1_0784_CZ1898_062.ab
1;
Oapa1_0479_CZ1835_090.ab
1;

arctic.fasta.screen.Contig62	94	NA	754	0	-	0	-	Oapa1_1540_AM23_091.ab1; Oapa1_0073_HFY327_055.ab1; Oapa1_1056_CZ1898_050.ab1; Oapa1_1048_CZ1898_049.ab1; Oapa1_0993_CZ1898_001.ab1; Oapa1_0763_CZ1898_044.ab1; Oapa1_0759_CZ1898_035.ab1; Oapa1_0651_CZ1855_039.ab1; Oapa1_0836_CZ1898_009.ab1; Oapa1_0087_HFY327_077.ab1; Oapa1_0024_HFY327_010.ab1; Oapa1_1395_EY0016_053.ab1; Oapa1_0074_HFY327_053.ab1; Oapa1_0807_CZ1898_076.ab1; Oapa1_0935_CZ1898_075.ab1; Oapa1_1260_EY0016_046.ab1; Oapa1_0223_HFY329_066.ab1	Oapa1_0015_HFY327_007.ab1; Oapa1_0259_CZ1835_006.ab	GW696963 GW69704

1;	6 -
Oapa1_0886_LS228_010M13	GW69687
F.ab1;	1 -
Oapa1_0875_LS228_015.ab1;	GW69703
Oapa1_0912_LS228_037.ab1;	0 -
Oapa1_0207_HFY329_075.ab	GW69703
1;	3
Oapa1_0935_LS228_056.ab1;	GW69740
Oapa1_0217_HFY329_078.ab	0
1;	GW69736
Oapa1_1167_CZ1898_077.ab	2 -
1;	GW69740
Oapa1_1088_CZ1898_094.ab	5 - -
1;	GW69755
Oapa1_0216_HFY329_080.ab	3
1;	GW69718
Oapa1_1176_CZ1898_080.ab	1 -
1;	GW69699
Oapa1_1459_AM23_008.ab1;	2 -
Oapa1_0592_CZ1855_093.ab	GW69721
1;	7
Oapa1_1534_AM23_072.ab1;	GW69754
Oapa1_0601_CZ1855_096.ab	2 -
1;	GW69696
Oapa1_0990_CZ1898_005.ab	0
1;	GW69693
Oapa1_0092_HFY327_067.ab	3 - - - - -
1;	GW69693
Oapa1_0424_CZ1835_038.ab	6 -
1;	GW69690
Oapa1_0679_CZ1855_050.ab	9 - - -
1;	GW69752
Oapa1_1508_AM23_062.ab1;	4
Oapa1_1446_AM23_011.ab1;	GW69748
Oapa1_0967_LS228_088.ab1;	4 - - - - - -
Oapa1_0936_LS228_054.ab1;	- - - - - -

Oapa1_1379_EY0016_046.ab1; Oapa1_0420_CZ1835_042.ab1; Oapa1_1092_CZ1898_086.ab1; Oapa1_0353_CZ1835_096.ab1; Oapa1_1041_CZ1898_063.ab1; Oapa1_0940_0.4uLdGTP_JF1427_079.ab1; Oapa1_1026_CZ1898_041.ab1; Oapa1_0899_0.4uLdGTP_JF1427_063.ab1; Oapa1_0788_CZ1898_058.ab1; Oapa1_0838_CZ1898_005.ab1; Oapa1_0999_CZ1898_008.ab1; Oapa1_1461_AM23_004.ab1; Oapa1_1367_EY0016_018.ab1; Oapa1_0931_LS228_064.ab1; Oapa1_1220_EY0016_010.ab1; Oapa1_1012_CZ1898_019.ab1; Oapa1_0931_CZ1898_050.ab1; Oapa1_1030_CZ1898_035.ab1; Oapa1_0816_CZ1898_091.ab1;	GW697310 GW697305 GW697398 - GW697295 GW697107 - - - - - - - GW697077 - - - - - - -

Oapa1_1021_CZ1898_020.ab1;
Oapa1_0706_CZ1855_085.ab1;
Oapa1_0513_CZ1855_021.ab1;
Oapa1_1230_EY0016_025.ab1;
Oapa1_0093_HFY327_065.ab1;
Oapa1_0875_JF1425_007N2R.ab1;
Oapa1_0964_CZ1898_093.ab1;
Oapa1_0940_CZ1898_067.ab1;
Oapa1_1164_CZ1898_052.ab1;
Oapa1_0918_LS228_042.ab1;
Oapa1_1115_CZ1898_029.ab1;
Oapa1_0672_CZ1855_062.ab1;
Oapa1_0488_CZ1855_013.ab1;
Oapa1_0461_CZ1835_076.ab1;
Oapa1_0351_CZ1835_083.ab1;
Oapa1_0265_CZ1835_031.ab1;
Oapa1_0181_HFY329_057.ab1;
Oapa1_0292_CZ1835_044.ab1;
Oapa1_1327_EY0016_015.ab

1;
Oapa1_1422_EY0016_068.ab
1;
Oapa1_1272_EY0016_057.ab
1;
Oapa1_1447_AM23_009.ab1;
Oapa1_1495_AM23_038.ab1;
Oapa1_1285_EY0016_079.ab
1;
Oapa1_0169_HFY329_033.ab
1;
Oapa1_0157_HFY329_026.ab
1;
Oapa1_0958_LS228_089.ab1;
Oapa1_1247_EY0016_047.ab
1;
Oapa1_1205_EY0016_013.ab
1;
Oapa1_1118_CZ1898_023.ab
1;
Oapa1_1541_AM23_089.ab1;
Oapa1_1004_CZ1898_029.ab
1;
Oapa1_1370_EY0016_043.ab
1;
Oapa1_0904_LS228_022.ab1;
Oapa1_0897_CZ1898_042.ab
1;
Oapa1_0385_CZ1835_027.ab
1;
Oapa1_0167_HFY329_037.ab
1;
Oapa1_1046_CZ1898_053.ab
1;
Oapa1_0645_CZ1855_018.ab
1;

arctic.fasta.screen.Contig60	79	hit to soybean genome	489	0	-	0	-	Oapa1_0062_HFY327_044.ab1; Oapa1_0964_LS228_094.ab1; Oapa1_0819_CZ1898_087.ab1; Oapa1_0889_CZ1898_048.ab1; Oapa1_0325_CZ1835_069.ab1; Oapa1_1326_EY0016_082.ab1; Oapa1_0099_HFY327_070.ab1; Oapa1_0877_CZ1898_045.ab1; Oapa1_1033_CZ1898_046.ab1; Oapa1_0039_HFY327_017.ab1; Oapa1_0889_LS228_004.ab1; Oapa1_0897_SB1906_091.ab1 Oapa1_0882_CZ1898_037.ab1; Oapa1_0882_CH0009_081.ab1; Oapa1_1263_EY0016_042.ab1; Oapa1_0882_JB981_039PCR_M13F.ab1; Oapa1_0091_HFY327_069.ab1; Oapa1_0280_CZ1835_018.ab1; Oapa1_1506_AM23_049.ab1; Oapa1_1187_CZ1898_091.ab1;	GW697291 GW696880 GW697440 GW696881 GW696991 GW697056 GW697541 GW69741

Oapa1_0882_betaine_1uLBDT_JF1427_045.ab1;	3 GW69687
Oapa1_0882_0.4uLdGTP_JF1427_095.ab1;	9 GW69687
Oapa1_0145_HFY329_029.ab1;	6 GW69700
Oapa1_0523_CZ1855_022.ab1;	7 GW69715
Oapa1_0212_HFY329_069.ab1;	9 GW69703
Oapa1_1265_EY0016_040.ab1;	2 GW69744
Oapa1_0669_CZ1855_051.ab1;	1 - GW69698
Oapa1_0063_HFY327_042.ab1;	2 GW69706
Oapa1_0302_CZ1835_061.ab1;	9 GW69728
Oapa1_0860_CZ1898_023.ab1;	1 GW69742
Oapa1_1224_EY0016_004.ab1;	4 GW69696
Oapa1_0027_HFY327_004.ab1;	8 GW69741
Oapa1_1199_CZ1898_088.ab1;	8 GW69749
Oapa1_1386_EY0016_036.ab1;	2 GW69695
Oapa1_0963_LS228_096.ab1;	7
Oapa1_0352_CZ1835_081.ab1;	GW69709 0
Oapa1_1085_CZ1898_081.ab1;	GW69736 1
Oapa1_0115_HFY327_096.ab1;	GW69699 5
Oapa1_1407_EY0016_079.ab	GW69750

1;	2
Oapa1_1433_EY0016_096.ab	GW69751
1;	4
Oapa1_1081_CZ1898_087.ab	GW69735
1;	9
Oapa1_0536_CZ1855_035.ab	GW69716
1;	4
Oapa1_0375_CZ1835_016.ab	GW69710
1;	0
Oapa1_0193_HFY329_060.ab	GW69702
1;	4
Oapa1_0107_HFY327_091.ab	GW69699
1;	3
Oapa1_0705_CZ1855_087.ab	GW69722
1;	8
Oapa1_1299_EY0016_074.ab	GW69745
1;	9
Oapa1_1226_EY0016_031.ab	GW69742
1;	5
Oapa01_0006_CZ1321_025.a	GW69756
b1;	1
Oapa1_0641_CZ1855_026.ab	GW69719
1;	9
Oapa1_0269_CZ1835_023.ab	GW69705
1;	0
Oapa1_1282_EY0016_054.ab	GW69745
1;	2
Oapa1_0923_LS228_063.ab1;	GW69692
Oapa1_0559_CZ1855_049.ab	3
1;	GW69716
Oapa1_0677_CZ1855_054.ab	9
1;	GW69721
Oapa1_0493_CZ1855_007.ab	6
1;	GW69714
Oapa1_0278_CZ1835_022.ab	5
1;	GW69705

Oapa1_1241_EY0016_028.ab1;	4 GW69743
Oapa1_1471_AM23_032.ab1;	0
Oapa1_0418_CZ1835_046.ab1;	GW69753 0
Oapa1_1512_AM23_056.ab1;	GW69712
Oapa1_0914_LS228_033.ab1;	1
Oapa1_0934_LS228_058.ab1;	GW69754
Oapa1_0922_LS228_034.ab1;	3
Oapa1_1537_AM23_066.ab1;	GW69691
Oapa1_1383_EY0016_040.ab1;	8 GW69693
Oapa1_0509_CZ1855_027.ab1;	2 GW69692
Oapa1_1089_CZ1898_092.ab1;	2 GW69755
Oapa1_0022_HFY327_012.ab1;	4 GW69749
Oapa1_0053_HFY327_043.ab1;	1 GW69715
Oapa1_0120_HFY327_092.ab1;	3 GW69736
Oapa1_0279_CZ1835_020.ab1;	3 GW69696
Oapa1_1457_AM23_012.ab1;	6
Oapa1_0996_CZ1898_014.ab1;	GW69697 8
Oapa1_0988_CZ1898_009.ab1;	GW69699 6
Oapa1_1436_EY0016_092.ab1;	GW69705 5
Oapa1_0799_CZ1898_073.ab1;	GW69752 2
Oapa1_0738_CZ1898_025.ab1;	GW69731 9
Oapa1_0721_CZ1898_009.ab	GW69731

								1; Oapa1_0760_CZ1898_033.ab1; Oapa1_0997_CZ1898_012.ab1; Oapa1_0723_CZ1898_005.ab1; Oapa1_0924_LS228_061.ab1; Oapa1_0963_CZ1898_095.ab1; Oapa1_0122_HFY327_088.ab1; Oapa1_0789_CZ1898_056.ab1; Oapa1_0965_LS228_092.ab1; Oapa1_1051_CZ1898_060.ab1; Oapa1_0882_1uLdGTP_JF1427_047.ab1; Oapa1_0934_CZ1898_077.ab1; Oapa1_0882_JF1425_088N2R.ab1	8 GW69751 5 GW69726 3 GW69724 4 GW69723 7 GW69725 1 GW69732 0 GW69723 8 GW69692 4 GW69730 9 GW69699 8 GW69725 7 GW69695 8 GW69733 9 GW69687 7 GW69730 3 -
arctic.fasta.screen.Contig55	9	very weak similarity to RAB18 COR protein	755	0	-	0	-	Oapa1_1341_EY0016_006.ab1; Oapa1_0613_CZ1855_007.ab1; Oapa1_0726_CZ1898_016.ab	GW69747 4 GW69718 6 GW69723

								1; Oapa1_0591_CZ1855_095.ab1; Oapa1_1172_CZ1898_071.ab1; Oapa1_1442_EY0016_084.ab1; Oapa1_0056_HFY327_037.ab1; Oapa1_0782_CZ1898_049.ab1; Oapa1_0252_CZ1835_003.ab1	9 GW697179 GW697402 - - GW697255 GW697043
arctic.fasta.screen.Contig52	8	NA	443	0	-	0	-	Oapa1_0783_CZ1898_064.ab1; Oapa1_1275_EY0016_051.ab1; Oapa1_1208_EY0016_009.ab1; Oapa1_0966_LS228_090.ab1; Oapa1_0048_HFY327_020.ab1; Oapa1_0966_CZ1898_091.ab1; Oapa1_0833_CZ1898_015.ab1; Oapa1_0678_CZ1855_052.ab1	GW697256 GW697446 - GW696959 GW696975 - GW697275 -
arctic.fasta.screen.Contig53	8	NA	390	0	-	0	-	Oapa1_0162_HFY329_047.ab1; Oapa1_0078_HFY327_062.ab1; Oapa1_0312_CZ1835_060.ab1; Oapa1_1520_AM23_075.ab1; Oapa1_1080_CZ1898_089.ab	GW697012 GW696987 - - GW697358 - - -

								1; Oapa1_0724_CZ1898_003.ab1; Oapa1_0290_CZ1835_048.ab1; Oapa1_0089_HFY327_073.ab1	
*arctic.fasta.screen.Contig48	5	NA	244	0	-	0	-	Oapa1_0751_CZ1898_018.ab1; Oapa1_1542_AM23_087.ab1; Oapa1_1414_EY0016_065.ab1; Oapa1_1296_EY0016_078.ab1; Oapa1_0308_CZ1835_051.ab1	GW69724 7 GW69755 6 GW69750 5 GW69745 7 GW69707 2
arctic.fasta.screen.Contig42	3	NA	459	0	-	0	-	Oapa1_0397_CZ1835_024.ab1; Oapa1_0076_HFY327_049.ab1; Oapa1_0537_CZ1855_033.ab1	GW69711 4 - GW69716 5
arctic.fasta.screen.Contig45	3	NA	360	0	-	0	-	Oapa1_0887_LS228_008.ab1; Oapa1_0880_LS228_005M13F.ab1; Oapa1_0880_JF1425_072N2R.ab1	GW69689 0 GW69687 5 GW69687 4
arctic.fasta.screen.Contig40	2	very weak similarity to ERD10	316	0	-	0	-	Oapa1_0906_LS228_018.ab1; Oapa1_0906_CZ1898_061.ab1	GW69691 4 -
arctic.fasta.screen.Contig11	1	NA	431	0	-	0	-	Oapa1_0296_CZ1835_036.ab1	GW69706 6
arctic.fasta.	1	NA	338	0	-	0	-	Oapa1_1501_AM23_059.ab1	GW69753

screen.Contig4									9
arctic.fasta.screen.Contig19	1	NA	283	0	-	0	-	Oapa1_0956_LS228_093.ab1	GW696952
arctic.fasta.screen.Contig22	1	NA	187	0	-	0	-	Oapa1_0926_LS228_057.ab1	-
Oapa1_0383_CZ1835_031.ab1	1	hit to soybean genome	502	0	-	0	-		GW697105
Oapa1_0041_HFY327_032.ab1	1	NA	500	0	-	0	-		GW696971
Oapa1_0271_CZ1835_019.ab1	1	NA	494	0	-	0	-		GW697052
Oapa1_0813_CZ1898_066.ab1	1	NA	249	0	-	0	-		GW697268
Oapa1_0983_CZ1898_084.ab1	1	NA	243	0	-	0	-		GW697316
Oapa1_0755_CZ1898_043.ab1	1	NA	155	0	-	0	-		-
Oapa1_0890_T12V_CH0009_018.ab1	1	NA	149	0	-	0			-
Oapa1_0069_HFY327_061.ab1	1	NA	118	0	-	0	-		-
Oapa1_0959_CZ1898_066.ab1	1	NA	110	0	-	0	-		-

Oapa1_0909_LS228_043.ab1	1	NA	107	0	-	0	-	-
Oapa1_0959_LS228_087.ab1	1	NA	107	0	-	0	-	-
Oapa1_0722_CZ1898_007.ab1	1	NA	106	0	-	0	-	-
Oapa1_0958_CZ1898_068.ab1	1	NA	105	0	-	0	-	-
Oapa1_0885_LS228_012M13F.ab1	1	NA	104	0	-	0	-	-
Oapa1_0246_CZ1835_013.ab1	1	NA	103	0	-	0	-	-
Oapa1_0814_CZ1898_095.ab1	1	NA	101	0	-	0	-	-
Oapa1_0154_HFY329_032.ab1	1	NA	98	0	-	0	-	-

A vertical line before the first column indicates that adjacent contigs have similar sequence.

* An asterisk indicates there is a similar contig in the temperate *Oxytropis* enriched library, and that the gene is a potential false positive.

8.4 Appendix 4: Supplementary Table S3.4 Annotation of the 368 genes (contigs and singlets) from a temperate-enriched *Oxytropis* subtracted plantlet library.

Temperate unique sequence name	Nb of ESTs in contigs	Gene putative identity	Length	Number of BLAST hits	Max e-value	Similarity mean	Number of GO terms	GO terms	Enzyme	Member sequences	GenBank Accession Numbers
		Energy									
		photosystems									
Otpt1_0074_HFY328_055.ab1	1	Chlorophyll a/b-binding protein; LHCA1; photosystem I	643	25	1.61E-83	92.28 %	9	P:photosynthesis, light harvesting; C:photosystem II; F:magnesium ion binding; P:protein-chromophore linkage; F:chlorophyll binding; C:integral to membrane; C:chloroplast; C:thylakoid; C:photosystem I			GW697586
temperate.fasta.screen.Contig110	5	Chlorophyll a b-binding protein; LHCA3 psi type III; 24kD light-harvesting protein of photosystem I	764	25	6.64E-91	86.76 %	9	P:photosynthesis, light harvesting; C:photosystem II; F:magnesium ion binding; P:protein-chromophore linkage; F:chlorophyll binding; C:integral to membrane; C:chloroplast; C:thylakoid; C:photosystem I		Otpt1_0276_HFY330_081.ab1; Otpt1_0784_CZ1855_064.ab1; Otpt1_0762_CZ1855_046.ab1; Otpt1_1002_CH0101_021.ab1; Otpt1_0331_CZ1835_028.ab1	GW697663 GW697846 GW697840 GW697925 GW697685
temperate.fasta.screen.Contig45	2	Chlorophyll a b-binding protein cp29; LHCB4; chloroplast protein encoded by nuclear genome	996	25	3.04E-146	89.32 %	5	C:plastoglobule; C:chloroplast envelope; P:photosynthesis, light harvesting; C:light-harvesting complex; C:chloroplast thylakoid membrane		Otpt1_0900_CH0101_036.ab1; Otpt1_0072_HFY328_059.ab1	GW697890 GW697584

temperate.fasta.screen.Contig99	4	Chlorophyll a-b binding protein P4; LHCI type III; chloroplast protein encoded by nuclear genome; Down-regulated by UV-B	1010	25	3.55E-134	87.28 %	9	P:photosynthesis, light harvesting; C:photosystem II; F:magnesium ion binding; P:protein-chromophore linkage; F:chlorophyll binding; C:integral to membrane; C:chloroplast; C:thylakoid; C:photosystem I	Otpt1_0144_HFY330_016.ab1; Otpt1_1259_CH0101_003.ab1; Otpt1_1466_AM21_012.ab1; Otpt1_1108_CH0101_088.ab1	GW697612 GW698029 GW698078 GW697965
Otpt1_1153_CH0101_041.ab1	1	Chlorophyll-a b binding protein; chloroplast protein encoded by nuclear genome	754	25	2.06E-129	97.24 %	9	P:photosynthesis, light harvesting; C:photosystem II; F:magnesium ion binding; P:protein-chromophore linkage; F:chlorophyll binding; C:integral to membrane; C:chloroplast; C:thylakoid; C:photosystem I		GW697992
Otpt1_1115_CH0101_011.ab1	1	Chlorophyll a b binding protein type II; chloroplast protein encoded by nuclear genome	840	25	1.30E-125	93.44 %	10	P:photosynthesis, light harvesting; C:plasma membrane; C:photosystem II; F:magnesium ion binding; P:protein-chromophore linkage; F:chlorophyll binding; C:integral to membrane; C:chloroplast; C:thylakoid; C:photosystem I		GW697970
Otpt1_0522_CZ1835_052.ab1	1	Chlorophyll a b binding protein	691	25	3.47E-61	96.76 %	9	P:photosynthesis, light harvesting; C:photosystem II; F:magnesium ion binding; P:protein-chromophore linkage; F:chlorophyll binding; C:integral to membrane; C:chloroplast; C:thylakoid; C:photosystem I		GW697750
temperate.fasta.screen.Contig119	8	Chlorophyll a b binding protein; LHCII type I CAB-AB80; Lhcb1; photosystem II	957	25	6.76E-140	94.80 %	9	P:photosynthesis, light harvesting; C:photosystem II; F:magnesium ion binding; P:protein-chromophore linkage; F:chlorophyll	Otpt1_0098_HFY328_065.ab1; Otpt1_1090_CH0101_068.ab1; Otpt1_0208_HFY330	GW697594 GW697956 GW697637 GW697643 GW697704

		type I; chloroplast protein encoded by nuclear genome						binding; C:integral to membrane; C:chloroplast; C:thylakoid; C:photosystem I	_034.ab1; Otpt1_0231_HFY330 _062.ab1; Otpt1_0395_CZ1835 _050.ab1; Otpt1_0595_CZ1855 _023.ab1; Otpt1_0937_CH0101 _069.ab1; Otpt1_0799_CZ1855 _069.ab1	GW697781 GW697904 GW697851
temperate.fasta.screen.Contig21	1	Chlorophyll a b binding protein	609	25	1.80E-78	97.84 %	8	P:photosynthesis, light harvesting; C:photosystem II; F:magnesium ion binding; P:protein-chromophore linkage; F:chlorophyll binding; C:integral to membrane; C:photosystem I; C:chloroplast thylakoid membrane	Otpt1_0042_HFY328 _028.ab1	GW697575
temperate.fasta.screen.Contig98	4	Chlorophyll a b binding protein type II; Cab-6; LHCB2.3; Photosystem II light harvesting complex gene 2.1; chloroplast protein encoded by nuclear genome;	1023	25	5.26E-133	93.44 %	9	P:photosynthesis, light harvesting; C:photosystem II; F:magnesium ion binding; P:protein-chromophore linkage; F:chlorophyll binding; C:integral to membrane; C:chloroplast; C:thylakoid; C:photosystem I	Otpt1_0044_HFY328 _024.ab1; Otpt1_1115_CH0101 _011.ab1; Otpt1_0165_HFY330 _021.ab1; Otpt1_0522_CZ1835 _052.ab1	GW697576 GW697970 GW697621 GW697750
temperate.fasta.screen.Contig96	3	Chlorophyll a b-binding protein; chloroplast protein encoded by nuclear genome	877	25	5.50E-114	86.84 %	4	C:membrane; C:thylakoid; C:chloroplast; P:photosynthesis, light harvesting	Otpt1_0110_HFY328 _070.ab1; Otpt1_0134_HFY330 _015.ab1; Otpt1_0285_HFY330 _086.ab1	GW697597 GW697608 GW697667

temperate.fasta.screen.Contig92	3	Chlorophyll a b-binding protein; chloroplast protein encoded by nuclear genome	688	25	1.13E-96	93.24 %	9	P:photosynthesis, light harvesting; C:photosystem II; F:magnesium ion binding; P:protein-chromophore linkage; F:chlorophyll binding; C:integral to membrane; C:chloroplast; C:thylakoid; C:photosystem I	Otpt1_0114_HFY328_066.ab1; Otpt1_0263_HFY330_074.ab1; Otpt1_0389_CZ1835_052.ab1	GW697598 GW697652 GW697703
temperate.fasta.screen.Contig49	2	Photosystem II 5 kD protein, chloroplast, also response to UV and wounding	513	25	2.82E-42	84.00 %	10	F:protein binding; C:stromule; F:carbonate dehydratase activity; P:carbon utilization; C:chloroplast; C:thylakoid; F:zinc ion binding; C:apoplast; P:one-carbon compound metabolic process; P:nitrogen compound metabolic process	Otpt1_0731_CZ1855_023.ab1; Otpt1_0211_HFY330_061.ab1	GW697830 GW697639
*Otpt1_0709_CZ1855_003.ab1	1	psaB; photosystem I p700 apoprotein a2	658	25	3.46E-121	99.20 %	14	F:electron carrier activity; F:iron ion binding; P:transport; P:oxidation reduction; F:magnesium ion binding; P:protein-chromophore linkage; F:chlorophyll binding; F:4 iron, 4 sulfur cluster binding; C:integral to membrane; C:chloroplast; C:thylakoid; C:photosystem I; P:photosynthesis; P:electron transport		GW697826
temperate.fasta.screen.Contig55	2	psaD; photosystem i subunit	773	25	4.74E-76	81.56 %	5	C:thylakoid; C:photosystem I reaction center; F:protein binding; C:chloroplast; P:photosynthesis	Otpt1_0273_HFY330_085.ab1; Otpt1_1204_CH0101_077.ab1	GW697661 GW698011
temperate.fasta.screen.Contig12	1	psaE family; reaction center subunit III	571	25	8.43E-37	78.92 %	5	C:thylakoid; F:catalytic activity; C:photosystem I reaction center; C:chloroplast;	Otpt1_1033_CH0101_033.ab1	GW697939

								P:photosynthesis		
temperate.fasta.screen.Contig56	2	psaE family; reaction center subunit iii	673	25	1.38E-35	79.96 %	5	C:thylakoid; F:catalytic activity; C:photosystem I reaction center; C:chloroplast; P:photosynthesis	Otpt1_1403_CH0101_050.ab1; Otpt1_1462_AM21_003.ab1	GW698054 GW698074
temperate.fasta.screen.Contig109	4	psaE family; Photosystem I reaction center subunit IV A, chloroplast protein encoded by nuclear genome	683	25	1.99E-29	87.40 %	5	C:thylakoid; F:catalytic activity; C:photosystem I reaction center; C:chloroplast; P:photosynthesis	Otpt1_1363_CH0101_045.ab1; Otpt1_0290_CZ1835_011.ab1; Otpt1_1174_CH0101_061.ab1; Otpt1_1119_CH0101_003.ab1	GW698038 GW697671 GW698002 GW697972
temperate.fasta.screen.Contig4	1	psaH; photosystem I protein	617	25	1.43E-62	86.00 %	5	C:integral to membrane; C:thylakoid; C:photosystem I reaction center; C:chloroplast; P:photosynthesis	Otpt1_0853_CH0101_014.ab1	GW697868
temperate.fasta.screen.Contig16	1	psaH; photosystem I protein	599	25	3.46E-55	82.60 %	5	C:integral to membrane; C:thylakoid; C:photosystem I reaction center; C:chloroplast; P:photosynthesis	Otpt1_0327_CZ1835_032.ab1	GW697683
temperate.fasta.screen.Contig25	1	psaK; photosystem I reaction center subunit x	732	25	9.34E-47	84.72 %	5	C:integral to membrane; C:photosystem I; C:thylakoid; C:chloroplast; P:photosynthesis	Otpt1_0313_CZ1835_004.ab1	GW697677
temperate.fasta.screen.Contig33	1	psaK; photosystem I reaction center subunit x	637	25	6.90E-47	85.56 %	5	C:integral to membrane; C:photosystem I; C:thylakoid; C:chloroplast; P:photosynthesis	Otpt1_0414_CZ1835_076.ab1	GW697710
temperate.fasta.screen.Contig83	3	psaL; photosystem I reaction subunit xi; chloroplast	817	25	4.00E-92	88.00 %	5	C:integral to membrane; C:thylakoid; C:photosystem I reaction center;	Otpt1_0793_CZ1855_079.ab1; Otpt1_1114_CH0101	GW697848 GW697969 GW697949

		protein encoded by nuclear genome					C:chloroplast; P:photosynthesis		_013.ab1; Otpt1_1070_CH0101 _054.ab1	
temperate.fasta.screen.Contig120	9	psaN; photosystem i reaction centre subunit precursor; calmodulin binding; unknown function ; chloroplast protein encoded by nuclear genome	691	25	7.23E-59	81.56 %	5	C:photosystem I; C:thylakoid membrane; F:protein binding; C:chloroplast; P:photosynthesis		Otpt1_1519_AM21_0 36.ab1; Otpt1_1457_AM21_0 13.ab1; Otpt1_0665_CZ1855 _069.ab1; Otpt1_1113_CH0101 _015.ab1; Otpt1_0129_HFY328 _090.ab1; Otpt1_0152_HFY330 _004.ab1; Otpt1_0077_HFY328 _051.ab1; Otpt1_0967_CH0101 _090.ab1; Otpt1_0130_HFY328 _088.ab1	GW698099 GW698072 GW697803 GW697968 GW697603 GW697614 GW697588 GW697914 GW697604
temperate.fasta.screen.Contig88	3	psaO, membrane protein 16kda; thylakoid; chloroplast protein encoded by nuclear genome	707	25	2.66E-56	79.24 %	2	C:thylakoid; C:plastid		Otpt1_0623_CZ1855 _046.ab1; Otpt1_0631_CZ1855 _036.ab1; Otpt1_1110_CH0101 _084.ab1	GW697792 GW697794 GW697967
Otpt1_1023_CH0101_018.ab1	1	psbM; photosystem II m protein	345	1	3.45E-07	75.00 %	0		-		GW697931
temperate.fasta.screen.Contig47	2	psbO; 33 kda subunit of oxygen evolving system of photosystem II thylakoid membrane protein; Mn stability; chloroplast protein	551	25	4.04E-54	94.64 %	6	F:manganese ion binding; C:extrinsic to membrane; F:calcium ion binding; P:photosystem II stabilization; C:chloroplast; C:oxygen evolving complex		Otpt1_0772_CZ1855 _061.ab1; Otpt1_0543_CZ1835 _066.ab1	GW697843 GW697763

temperate.fasta.screen.Contig115	6	psbP; 23 kda subunit of oxygen evolving system of photosystem II thylakoid membrane protein; chloroplast protein encoded by nuclear genome	911	25	5.89E-106	84.16 %	7	C:chloroplast envelope; C:extrinsic to membrane; F:calcium ion binding; C:stromule; C:apoplast; P:photosynthesis; C:oxygen evolving complex	Otpt1_0571_CZ1855_005.ab1; Otpt1_1392_CH0101_053.ab1; Otpt1_1013_CH0101_028.ab1; Otpt1_1505_AM21_033.ab1; Otpt1_0361_CZ1835_034.ab1; Otpt1_1091_CH0101_066.ab1	GW697776 GW698049 GW697928 GW698094 GW697693 GW697957
temperate.fasta.screen.Contig1	1	psbP; 23 kda subunit of oxygen evolving system of photosystem II thylakoid membrane protein; chloroplast protein encoded by nuclear genome	659	25	2.50E-95	78.76 %	7	C:chloroplast envelope; C:extrinsic to membrane; F:calcium ion binding; C:stromule; C:apoplast; P:photosynthesis; C:oxygen evolving complex	Otpt1_0365_CZ1835_057.ab1	GW697697
Otpt1_0329_CZ1835_030.ab1	1	psbQ; oxygen evolving enhancer protein; calcium binding; light reaction; chloroplast protein encoded by nuclear genome	661	25	3.94E-64	70.92 %	6	C:chloroplast thylakoid lumen; C:extrinsic to membrane; F:calcium ion binding; P:photosynthesis; C:chloroplast thylakoid membrane; C:oxygen evolving complex		GW697684
Otpt1_0988_CH0101_010.ab1	1	psbQ; oxygen-evolving enhancer protein 3 precursor; chloroplast protein encoded by nuclear genome; only weakely similar to Otpt1_0329	767	25	3.13E-64	57.52 %	4	C:chloroplast thylakoid lumen; P:photosynthesis, light reaction; C:chloroplast thylakoid membrane; C:oxygen evolving complex		GW697921

temperate.fasta.screen.Contig114	6	psbQ; oxygen-evolving enhancer protein 3 precursor; calcium binding; light reaction; chloroplast protein encoded by nuclear genome; only weakly similar to Otpt1_0329 or Otpt1_0988	829	25	6.84E-63	85.44 %	5	C:extrinsic to membrane; F:calcium ion binding; C:chloroplast; P:photosynthesis; C:oxygen evolving complex		Otpt1_0741_CZ1855_028.ab1; Otpt1_0906_CH0101_063.ab1; Otpt1_1116_AM21_065.ab1; Otpt1_1124_CH0101_010.ab1; Otpt1_1394_CH0101_049.ab1; Otpt1_0843_CH0101_011.ab1	GW697834 GW697892 GW698071 GW697976 GW698050 GW697865
temperate.fasta.screen.Contig117	7	psbR; photosystem II polypeptide 10 kda ; chloroplast protein encoded by nuclear genome	710	25	2.21E-50	86.36 %	4	C:oxygen evolving complex; C:thylakoid membrane; C:chloroplast; P:photosynthesis		Otpt1_0146_HFY330_012.ab1; Otpt1_1379_CH0101_044.ab1; Otpt1_0184_HFY330_043.ab1; Otpt1_0425_CZ1835_091.ab1; Otpt1_1414_CH0101_067.ab1; Otpt1_1083_CH0101_080.ab1; Otpt1_0249_HFY330_069.ab1	GW697613 GW698044 GW697630 GW697714 GW698058 GW697954 GW697649
temperate.fasta.screen.Contig27	1	psbR chloroplast photosystem II 10 kDa protein; chloroplast protein encoded by nuclear genome	270	0			0		-	Otpt1_1399_CH0101_058.ab1	GW698052
temperate.fasta.screen.Contig84	3	psbW; photosystem II reaction center chloroplast precursor	641	25	8.86E-34	73.80 %	2	C:chloroplast; C:photosystem II		Otpt1_0063_HFY328_044.ab1; Otpt1_1146_CH0101_024.ab1; Otpt1_0903_CH0101_034.ab1	GW697579 GW697987 GW697891

temperate.fasta.screen.Contig67	2	psbW; photosystem ii reaction center chloroplast precursor	641	25	3.05E-26	77.28 %	5	C:integral to membrane; C:thylakoid; C:chloroplast; P:photosynthesis; C:photosystem II		Otpt1_0943_CH0101_076.ab1; Otpt1_0656_CZ1855_050.ab1	GW697908 GW697800
temperate.fasta.screen.Contig94	3	psbX; photosystem II ; ultraviolet-b-repressible protein	623	25	8.80E-28	76.68 %	2	P:photosynthesis; C:photosystem II		Otpt1_0979_CH0101_009.ab1; Otpt1_0809_CZ1855_072.ab1; Otpt1_0789_CZ1855_056.ab1	GW697917 GW697854 GW697847
Otpt1_0885_CH0101_047.ab1	1	psbY; Photosystem II protein	596	25	6.07E-28	78.60 %	1	C:membrane	-		GW697880
		Accessory proteins									
Otpt1_0381_CZ1835_058.ab1	1	Magnesium-protoporphyrin IX monomethyl ester [oxidative] cyclase; leucine zipper protein with DNA binding; chloroplast	215	12	2.54E-10	73.08 %	2	F:metal ion binding; P:metabolic process			GW697700
temperate.fasta.screen.Contig76	2	petC; cytochrome b6-f complex (rieske iron-sulfur protein) (plastohydroquinone:plastocyanin oxidoreductase iron-sulfur protein); chloroplast protein encoded by nuclear genome	939	25	1.88E-94	83.68 %	14	F:plastoquinol-plastocyanin reductase activity; F:ubiquinol-cytochrome-c reductase activity; C:plasma membrane; F:electron carrier activity; F:iron ion binding; P:oxidation reduction; F:2 iron, 2 sulfur cluster binding; C:integral to membrane; C:chloroplast; P:electron transport; C:cytochrome b6f complex; P:oxidative phosphorylation; P:proton transport;	EC:1.10.99.1, EC:1.10.2.2,	Otpt1_0463_CZ1835_002.ab1; Otpt1_0554_CZ1835_081.ab1	GW697724 GW697770

Otpt1_0855_CH0101_010.ab1	1	petC; cytochrome b6-f complex (rieske iron-sulfur protein) (plastohydroquinone:plastocyanin oxidoreductase iron-sulfur protein); chloroplast protein encoded by nuclear genome	754	25	6.46E-99	84.96 %	14	F:plastoquinol-plastocyanin reductase activity; F:ubiquinol-cytochrome-c reductase activity; C:plasma membrane; F:electron carrier activity; F:iron ion binding; P:oxidation reduction; F:2 iron, 2 sulfur cluster binding; C:integral to membrane; C:chloroplast; P:electron transport; C:cytochrome b6f complex; P:oxidative phosphorylation; P:proton transport;	EC:1.10.99.1, EC:1.10.2.2,		GW697870
temperate.fasta.screen.Contig30	1	petC; rieske iron-sulphur protein precursor, cytochrome C in photosystem II; chloroplast protein encoded by chloroplast genome	603	25	1.06E-15	89.96 %	13	F:plastoquinol-plastocyanin reductase activity; F:ubiquinol-cytochrome-c reductase activity; F:electron carrier activity; F:iron ion binding; P:oxidation reduction; F:2 iron, 2 sulfur cluster binding; C:integral to membrane; C:chloroplast; P:electron transport; C:cytochrome b6f complex; P:oxidative phosphorylation; P:proton transport;	-	Otpt1_0686_CZ1855_085.ab1	GW697811
temperate.fasta.screen.Contig108	4	petE; plastocyanin –thylakoid lumen; copper ion binding; chloroplast protein encoded by nuclear genome	691	25	5.77E-45	87.72 %	8	F:copper ion binding; C:thylakoid lumen; F:electron carrier activity; P:transport; C:stromule; C:chloroplast; C:membrane; P:electron transport		Otpt1_0590_CZ1855_031.ab1; Otpt1_0335_CZ1835_022.ab1; Otpt1_0070_HFY328_063.ab1; Otpt1_0982_CH0101_003.ab1	GW697778 GW697686 GW697582 GW697919
Otpt1_1289_CH0101_030.ab1	1	petM; cytochrome b6f complex subunit	546	22	1.28E-28	74.95 %	1	C:chloroplast thylakoid membrane	EC:1.10.99.1,		GW698034

Otpt1_0907_CH0101_061.ab1	1	Thioredoxin h; act on sulfur group of donors	563	25	8.57E-39	83.24 %	7	P:cell redox homeostasis; C:cytoplasm; F:electron carrier activity; F:protein disulfide oxidoreductase activity; P:glycerol ether metabolic process; P:transport; P:electron transport			GW697893
Otpt1_0911_CH0101_055.ab1	1	Thioredoxin m; chloroplast precursor	416	25	9.71E-31	93.16 %	8	F:electron carrier activity; F:protein binding; P:cell redox homeostasis; P:glycerol ether metabolic process; P:transport; C:chloroplast; F:protein disulfide oxidoreductase activity; P:electron transport			GW697894
		Electron transport									
temperate.fasta.screen.Contig77	2	Chloroplast ferredoxin I	770	25	1.86E-48	82.24 %	10	P:ferredoxin metabolic process; F:ferrous iron binding; F:electron carrier activity; F:protein binding; P:transport; C:stromule; F:2 iron, 2 sulfur cluster binding; C:chloroplast; P:electron transport chain; P:electron transport		Otpt1_1178_CH0101_055.ab1; Otpt1_1132_CH0101_027.ab1	GW698004 GW697980
Otpt1_0188_HFY330_037.ab1	1	Chloroplast ferredoxin-nadp	586	25	5.15E-93	83.96 %	10	F:NADP binding; F:electron carrier activity; P:transport; P:oxidation reduction; C:thylakoid membrane; F:ferredoxin-NADP+ reductase activity; F:FAD binding; C:chloroplast; P:photosynthesis; P:electron transport	EC:1.18.1.2,		GW697632

Otpt1_1446_CH0101_092.ab1	1	Chloroplast ferredoxin-nadp+ reductase; very weak similiar similar to Otpt1_0188	660	25	1.87E-90	87.96 %	13	C:chloroplast envelope; F:NADP binding; F:electron carrier activity; F:protein binding; P:transport; P:oxidation reduction; C:stromule; C:thylakoid membrane; F:ferredoxin-NADP+ reductase activity; F:FAD binding; C:apoplast; P:photosynthesis; P:electron transport	EC:1.18.1.2,		GW698065
temperate.fasta.screen.Contig63	2	ferredoxin NADP+ reductase; chloroplast	458	25	5.17E-32	96.60 %	13	C:chloroplast envelope; F:NADP binding; F:electron carrier activity; F:protein binding; P:transport; P:oxidation reduction; C:stromule; C:thylakoid membrane; F:ferredoxin-NADP+ reductase activity; F:FAD binding; C:apoplast; P:photosynthesis; P:electron transport	EC:1.18.1.2,	Otpt1_0322_CZ1835_023.ab1; Otpt1_0465_CZ1835_029.ab1	GW697680 GW697725
temperate.fasta.screen.Contig116	6	Chloroplast precursor	719	25	3.74E-53	81.92 %	10	C:chloroplast thylakoid lumen; F:copper ion binding; F:DNA topoisomerase (ATP-hydrolyzing) activity; F:electron carrier activity; P:transport; C:DNA topoisomerase complex (ATP-hydrolyzing); C:stromule; P:DNA topological change; C:membrane; P:electron transport		Otpt1_0239_HFY330_050.ab1; Otpt1_0513_CZ1835_049.ab1; Otpt1_1021_CH0101_020.ab1; Otpt1_0065_HFY328_040.ab1; Otpt1_0645_CZ1855_049.ab1; Otpt1_0545_CZ1835_093.ab1	GW697647 GW697746 GW697930 GW697580 GW697798 GW697765
Otpt1_1490_AM21_022.ab1	1	Desaturase-like protein; chlorophyll biosynthesis; magnesium-	500	25	8.96E-59	95.36 %	7	F:iron ion binding; F:magnesium-protoporphyrin IX monomethyl ester (oxidative) cyclase activity; P:oxidation reduction;	EC:1.14.13.81,		GW698089

		protoporphyrin IX monomethyl ester					P:chlorophyll biosynthetic process; C:chloroplast; C:membrane; P:photosynthesis				
Otpt1_0915_CH0101_051.ab1	1	Magnesium-chelatase subunit; mg-protoporphyrin ix chelatase; chlorophyll biosynthesis; also nucleoside-triphosphatase activity	607	25	2.78E-39	92.96 %	7	F:magnesium chelatase activity; P:chlorophyll biosynthetic process; C:stromule; F:ATP binding; F:nucleoside-triphosphatase activity; P:photosynthesis; C:magnesium chelatase complex	EC:6.6.1.1, EC:3.6.1.15,		GW697895
		Proton transport									
Otpt1_1109_CH0101_086.ab1	1	ATP synthase (gamma subunit)	722	25	1.59E-89	85.08 %	7	C:proton-transporting ATP synthase complex, catalytic core F(1); F:hydrogen ion transporting ATP synthase activity, rotational mechanism; P:ATP synthesis coupled proton transport; F:metal ion binding; C:thylakoid; F:hydrogen ion transporting ATPase activity, rotational mechanism; C:chloroplast	EC:3.6.3.14,		GW697966
		Carbon fixation									
temperate.fasta.screen.Contig118	8	Ribulose-bisphosphate carboxylase small subunit	711	25	4.75E-85	85.68 %	7	P:reductive pentose-phosphate cycle; F:ribulose-bisphosphate carboxylase activity; F:monooxygenase activity; P:photorespiration; P:oxidation reduction; C:chloroplast ribulose bisphosphate carboxylase	EC:4.1.1.39,	Otpt1_0212_HFY330_059.ab1; Otpt1_0156_HFY330_031.ab1; Otpt1_0491_CZ1835_037.ab1; Otpt1_1097_CH0101_085.ab1;	GW697640 GW697616 - GW697960 GW697979 GW697787 GW697627 GW697832

								complex; P:glyoxylate metabolic process		Otpt1_1131_CH0101_029.ab1; Otpt1_0606_CZ1855_022.ab1; Otpt1_0181_HFY330_018.ab1; Otpt1_0734_CZ1855_017.ab1	
temperate.fasta.screen.Contig90	3	Ribulose-bisphosphate carboxylase small subunit	704	25	5.10E-84	85.44 %	7	P:reductive pentose-phosphate cycle; F:ribulose-bisphosphate carboxylase activity; F:monooxygenase activity; P:photorespiration; P:oxidation reduction; C:chloroplast ribulose bisphosphate carboxylase complex; P:glyoxylate metabolic process	EC:4.1.1.39,	Otpt1_0289_CZ1835_013.ab1; Otpt1_1508_AM21_048.ab1; Otpt1_1218_CH0101_070.ab1	GW697670 GW698095 GW698017
temperate.fasta.screen.Contig20	1	Ribulose-bisphosphate carboxylase small subunit	660	25	3.21E-82	85.80 %	7	P:reductive pentose-phosphate cycle; F:ribulose-bisphosphate carboxylase activity; F:monooxygenase activity; P:photorespiration; P:oxidation reduction; C:chloroplast ribulose bisphosphate carboxylase complex; P:glyoxylate metabolic process	EC:4.1.1.39,	Otpt1_0527_CZ1835_075.ab1	GW697753
temperate.fasta.screen.Contig57	2	Rubisco activase	690	25	8.39E-124	96.24 %	5	F:ribulose-bisphosphate carboxylase activity; F:ATP binding; C:chloroplast ribulose bisphosphate carboxylase complex; P:carbon utilization by fixation of carbon dioxide; P:glyoxylate metabolic process	EC:4.1.1.39,	Otpt1_0277_HFY330_096.ab1; Otpt1_1129_CH0101_002.ab1	GW697664 GW697977

Otpt1_0688_CZ1855_083.ab1	1	Ribulose bisphosphate carboxylase oxygenase activase large protein isoform; chloroplast	337	25	8.26E-22	81.96 %	5	F:ribulose-bisphosphate carboxylase activity; F:ATP binding; C:chloroplast ribulose bisphosphate carboxylase complex; P:carbon utilization by fixation of carbon dioxide; P:glyoxylate metabolic process	EC:4.1.1.39,	GW697812
Otpt1_1018_CH0101_022.ab1	1	Ribulose bisphosphate carboxylase oxygenase activase large protein isoform; rubisco activase; chloroplast	512	25	9.05E-25	80.12 %	11	F:ribulose-bisphosphate carboxylase activity; F:protein binding; C:cell wall; C:stromule; F:ATP binding; C:thylakoid; C:nucleus; C:apoplast; C:chloroplast ribulose bisphosphate carboxylase complex; P:carbon utilization by fixation of carbon dioxide; P:glyoxylate metabolic process	EC:4.1.1.39,	GW697929
Otpt1_1523_AM21_061.ab1	1	Oxydative phosphorylation Ubiquinol-cytochrome c reductase-like protein	614	25	6.78E-49	88.12 %	8	C:mitochondrial respiratory chain complex III; F:ubiquinol-cytochrome-c reductase activity; P:mitochondrial electron transport, ubiquinol to cytochrome c; F:electron transporter, transferring electrons within CoQH2-cytochrome c reductase complex activity; C:plastid; P:electron transport; P:proton transport;	EC:1.10.2.2,	GW698101

Otpt1_0923_CH0101_058.ab1	1	Vacuolar proton pump subunit e (v-atpase subunit e) (-1); membranes; proton transport	759	25	4.12E-101	89.16 %	7	C:cytosolic ribosome; C:vacuole; C:proton-transporting two-sector ATPase complex; C:plasma membrane; P:proton transport; F:hydrogen ion transporting ATPase activity, rotational mechanism; P:oxidative phosphorylation	EC:3.6.3.14,	GW697899
		Carbohydrates								
		Glycolysis and gluconoegenesis								
Otpt1_1416_CH0101_065.ab1	1	Fructose-bisphosphate aldolase; gluconeogenesis, pentose phosphate	613	25	2.76E-42	86.76 %	10	C:cytoplasm; F:protein binding; P:glycolysis; F:fructose-bisphosphate aldolase activity; P:fructose metabolic process; P:mannose metabolic process; P:inositol metabolic process; P:gluconeogenesis; P:pentose-phosphate shunt; P:carbon utilization	EC:4.1.2.13,	GW698059
Otpt1_0939_CH0101_065.ab1	1	Glyceraldehyde-3-phosphate dehydrogenase b subunit; chloroplast; pentose phosphate; response to light and sucrose	695	25	1.27E-95	94.40 %	11	F:glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) activity; F:NAD binding; P:reductive pentose-phosphate cycle; F:glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating) activity; P:oxidation reduction; P:glycolysis; C:chloroplast; P:response to cadmium ion; C:stromule; C:apoplast; P:gluconeogenesis	EC:1.2.1.12, EC:1.2.1.13,	GW697906

Otpt1_1310_CH0101_022.ab1	1	Malate dehydrogenase; chloroplast; Citrate cycle (TCA cycle) and Glyoxylate and dicarboxylate; Reductive carboxylate cycle (CO2 fixation) and pyruvate metabolism	496	25	3.11E-56	83.44 %	11	F:L-malate dehydrogenase activity; F:binding; P:plant-type hypersensitive response; P:response to cold; P:oxidation reduction; P:malate metabolic process; P:glycolysis; C:mitochondrial matrix; P:pyruvate metabolic process; P:reductive tricarboxylic acid cycle; P:glyoxylate metabolic process	EC:1.1.1.37,	GW698035
Otpt1_1032_CH0101_035.ab1	1	Triosephosphate isomerase; Carbon fixation; fructose and mannose; glucose and gluconeogenesis; inositol	726	25	1.50E-81	86.20 %	13	P:reductive pentose-phosphate cycle; F:triose-phosphate isomerase activity; F:protein binding; C:stromule; C:chloroplast; C:thylakoid; C:apoplast; P:fructose metabolic process; P:mannose metabolic process; P:inositol metabolic process; P:gluconeogenesis; P:glycolysis; P:glycerolipid metabolic process	EC:5.3.1.1,	GW697938
Otpt1_0267_HFY330_066.ab1	1	Triosephosphate isomerase; Carbon fixation; fructose and mannose; glucose and gluconeogenesis; inositol	680	25	1.87E-96	89.04 %	15	P:fatty acid biosynthetic process; F:acyltransferase activity; P:pentose-phosphate shunt; F:tetrahydroberberine oxidase activity; F:triose-phosphate isomerase activity; P:glycolysis; P:gluconeogenesis; C:cytoplasm; P:acyl-carrier-protein biosynthetic process; P:alkaloid biosynthetic process; P:fructose metabolic process; P:mannose metabolic process; P:inositol metabolic process; P:carbon	EC:1.3.3.8, EC:5.3.1.1,	GW697656

								utilization; P:glycerolipid metabolic process			
Otpt1_1004_CH0101_019.ab1	1	Phosphoglycerate kinase cytosolic; Glycolysis and Gluconeogenesis; Carbon fixation	542	25	1.22E-76	92.96 %	14	C:cell wall; F:phosphoglycerate kinase activity; C:thylakoid; F:ATP binding; P:glycolysis; C:chloroplast; P:response to cadmium ion; C:nucleus; C:plasma membrane; C:stromule; C:apoplast; P:gluconeogenesis; P:carbon utilization; P:phosphorylation	EC:2.7.2.3,		GW697926
Otpt1_0888_CH0101_041.ab1	1	Chloroplast fructose-bisphosphatase II	521	25	1.29E-53	85.68 %	9	F:fructose-bisphosphatase activity; P:response to cold; C:plastid; P:fructose metabolic process; P:mannose metabolic process; P:gluconeogenesis; P:glycolysis; P:pentose-phosphate shunt; P:carbon utilization			GW697882
temperate.fasta.screen.Contig37	2	Other carbohydrate metabolism Glucosyltransferase-like protein; mannane and cellulose synthase	273	25	9.16E-29	88.24 %	2	C:membrane; F:transferase activity	-	Otpt1_1192_CH0101_060.ab1; Otpt1_1271_CH0101_031.ab1	GW698006 GW698032
Otpt1_0269_HFY330_09	1	Basic chitinase; also amino-sugar	663	25	3.20E-64	87.44 %	3	F:chitinase activity; P:chitin catabolic process; P:cell wall	EC:3.2.1.14,		GW697657

3.ab1		metabolism (chitin and cell wall catabolism)						catabolic process			
Otpt1_1517_AM21_038.ab1	1	Beta-amylase; starch and sucrose	502	3	3.09E-22	62.67 %	2	P:carbohydrate metabolic process; F:hydrolase activity, acting on glycosyl bonds			GW698098
Otpt1_1087_CH0101_074.ab1	1	Chitinase related; class v; glycosyl hydrolase (chitin and cell wall catabolism)	476	25	6.27E-70	60.60 %	2	F:hydrolase activity, hydrolyzing O-glycosyl compounds; P:carbohydrate metabolic process	-		GW697955
Otpt1_0531_CZ1835_067.ab1	1	Beta 1-3 glucanase	698	25	3.02E-36	74.92 %	2	C:cytoplasmic membrane-bounded vesicle; F:catalytic activity	-		GW697757
temperate.fasta.screen.Contig86	3	Invertase pectin methylesterase inhibitor family protein; carbohydrate cell wall extension	853	25	1.02E-32	63.32 %	1	F:pectinesterase inhibitor activity	-	Otpt1_1452_CH0101_086.ab1; Otpt1_1535_AM21_060.ab1; Otpt1_1159_CH0101_048.ab1	GW698068 GW698106 GW697995
Otpt1_0600_CZ1855_017.ab1	1	UDP-galactose 4-epimerase; galactose and nucleotide sugar metabolism; stress response	427	25	7.26E-50	86.04 %	6	C:mitochondrion; F:coenzyme binding; F:UDP-glucose 4-epimerase activity; P:galactose metabolic process; P:nucleotide metabolic process; P:nucleotide-sugar metabolic process	EC:5.1.3.2,		GW697784
Otpt1_1377_CH0101_048.ab1	1	Xyloglucan endotransglycosylase	570	13	1.31E-13	75.00 %	5	F:hydrolase activity, hydrolyzing O-glycosyl compounds; C:apoplast; C:cell wall; P:glucan metabolic process; F:xyloglucan:xyloglucosyl transferase activity	EC:3.2.1, EC:2.4.1.207,		GW698043

Otpt1_0364_CZ1835_059.ab1	1	glycosyl hydrolase family protein 17; plasma membrane	455	25	6.54E-43	79.44 %	1	C:cytoplasmic membrane-bounded vesicle			GW697696
temperate.fasta.screen.Contig62	2	glycosyl hydrolase family 38	288	25	1.76E-27	84.56 %	0		-	Otpt1_0320_CZ1835_027.ab1; Otpt1_0320_AM21_092.ab1	GW697679 GW698070
Otpt1_0401_CZ1835_075.ab1	1	Glycoside hydrolase, family 5; ChaC-like protein	593	25	2.32E-64	88.80 %	2	P:response to lead ion; P:response to cadmium ion	-		GW697707
		Nitrogen metabolism – Amino acids									
*temperate.fasta.screen.Contig24	1	Carbonic anhydrase 1 carbonate dehydratase zinc ion binding	593	25	2.35E-24	86.28 %	10	F:protein binding; C:stromule; F:carbonate dehydratase activity; P:carbon utilization; C:chloroplast; C:thylakoid; F:zinc ion binding; C:apoplast; P:one-carbon compound metabolic process; P:nitrogen compound metabolic process	EC:4.2.1.1,	Otpt1_0832_CZ1855_092.ab1	GW697860
*temperate.fasta.screen.Contig42	2	Carbonic anhydrase	696	25	2.82E-42	84.00 %	10	F:protein binding; C:stromule; F:carbonate dehydratase activity; P:carbon utilization; C:chloroplast; C:thylakoid; F:zinc ion binding; C:apoplast; P:one-carbon compound metabolic process; P:nitrogen compound metabolic process	EC:4.2.1.1,	Otpt1_0166_HFY330_019.ab1; Otpt1_0536_CZ1835_076.ab1	GW697622 GW697759

Otpt1_1202_CH0101_050.ab1	1	Glutamine synthetase cytosolic; amino acid and Glycan biosynthesis metabolism	732	25	5.32E-34	96.80 %	5	C:cytoplasm; F:glutamate-ammonia ligase activity; P:nitrogen fixation; P:glutamine biosynthetic process; P:peptidoglycan biosynthetic process	EC:6.3.1.2,	GW698009
Otpt1_1079_CH0101_069.ab1	1	NADH-dependent glutamate synthase	426	0			0		-	GW697952
Otpt1_0159_HFY330_027.ab1	1	Nitrite reductase; NO forming (Nitrite -> Ammonia); sulphite and nitrite assimilation	665	25	5.16E-88	86.32 %	13	F:ferredoxin-nitrite reductase activity; F:electron carrier activity; P:nitrate assimilation; F:iron ion binding; F:protein binding; F:heme binding; P:transport; P:oxidation reduction; C:stromule; F:4 iron, 4 sulfur cluster binding; C:chloroplast; C:apoplast; P:electron transport	EC:1.7.7.1,	GW697618
		Protein degradation								
Otpt1_0960_CH0101_085.ab1	1	Gamma glutamyl hydrolase; urea cycle	641	25	4.15E-66	74.64 %	5	C:extracellular space; C:vacuole; F:gamma-glutamyl hydrolase activity; C:cell wall; P:folic acid biosynthetic process	-	GW697911
Otpt1_0011_HFY328_015.ab1	1	Glycine decarboxylase multi-enzyme h subunit; mitochondrion, photorespiration, other amino acid metabolism	629	25	2.28E-63	90.72 %	10	P:glycine catabolic process; F:glycine dehydrogenase (decarboxylating) activity; F:aminomethyltransferase activity; F:protein binding; F:lipoic acid binding; C:chloroplast; C:glycine cleavage complex; C:mitochondrion; P:L-serine metabolic process; P:threonine metabolic	EC:2.1.2.10,	GW697564

								process		
Otpt1_1205_CH0101_075.ab1	1	Glycine decarboxylase; system t protein; aminomethyltransferase; mitochondrion or cytoplasm	388	25	1.03E-64	93.88 %	7	P:glycine catabolic process; F:aminomethyltransferase activity; F:transaminase activity; C:glycine cleavage complex; C:mitochondrion; P:L-serine metabolic process; P:threonine metabolic process	EC:2.1.2.10, EC:2.6.1,	GW698012
Otpt1_0938_CH0101_067.ab1	1	Glycine dehydrogenase, system p-protein; mitochondrion, Glycine serine threonin pathway	749	25	1.14E-79	71.16 %	10	F:glycine dehydrogenase (decarboxylating) activity; F:pyridoxal phosphate binding; P:oxidation reduction; C:chloroplast; C:apoplast; C:glycine cleavage complex; P:glycine metabolic process; C:mitochondrion; P:L-serine metabolic process; P:threonine metabolic process	EC:1.4.4.2,	GW697905
Otpt1_0314_CZ1835_002.ab1	1	Serine carboxypeptidase	721	25	1.94E-41	78.84 %	5	F:serine carboxypeptidase activity; C:vacuole; C:cytoplasmic membrane-bounded vesicle; F:carboxypeptidase D activity; P:proteolysis	EC:3.4.16,	GW697678
		Protein folding or modification								

Otpt1_0071_HFY328_061.ab1	1	Calreticulin 3; chaperone-like; endoplasmic reticulum; Glycan bind calnexin, calcium binding	646	25	5.43E-79	92.84 %	6	P:protein folding; C:endomembrane system; C:cytoplasmic membrane-bounded vesicle; F:calcium ion binding; F:unfolded protein binding; C:endoplasmic reticulum			GW697583
Otpt1_1249_CH0101_084.ab1	1	Clathrin assembly protein ap19 homolog; trans-Golgi network	712	25	4.45E-83	95.52 %	9	P:vesicle-mediated transport; F:protein transporter activity; C:mitochondrion; C:clathrin coat of trans-Golgi network vesicle; C:plastid; C:clathrin adaptor complex; F:ATP binding; F:protein binding; P:intracellular protein transport			GW698025
temperate.fasta.screen.Contig38	2	Endoplasmic reticulum retrieval protein 1A; Golgi to ER	644	25	4.85E-80	87.76 %	1	C:membrane		Otpt1_0525_CZ1835_079.ab1; Otpt1_1143_CH0101_030.ab1	GW697751 GW697986
Otpt1_1246_CH0101_090.ab1	1	EXO70 protein B1; protein binding for exocytosis	561	25	1.41E-41	69.20 %	1	C:plastid			GW698023
Otpt1_1536_AM21_058.ab1	1	NSF n-ethylmaleimide sensitive factor; s atpases; chaperone-like, assemble protein complexes; ATP-binding	592	25	6.19E-25	84.52 %	2	F:ATP binding; F:nucleoside-triphosphatase activity	-		GW698107
Otpt1_1027_CH0101_045.ab1	1	Peptidyl-prolyl cis-trans isomerase PPIC-type family protein; similar to PIN1AT (parvulin 1At)	714	25	4.09E-60	83.96 %	2	F:peptidyl-prolyl cis-trans isomerase activity; P:protein folding	EC:5.2.1.8,		GW697933

Otpt1_0868_CH0101_017.ab1	1	Ubiquitin-conjugating enzyme	394	25	1.69E-54	94.00 %	5	F:ubiquitin-protein ligase activity; P:ubiquitin cycle; P:regulation of protein metabolic process; P:post-translational protein modification; P:protein ubiquitination	-	GW697875
Otpt1_0287_HFY330_082.ab1	1	ATG8C Autophagy-related ubiquitin-like modifier; with GABA-receptor-associated protein domain	629	25	1.38E-52	93.28 %	9	F:protein binding; C:autophagic vacuole; F:APG8-specific protease activity; F:APG8 activating enzyme activity; C:vacuolar lumen; F:APG8 ligase activity; C:mitochondrion; P:autophagy; P:proteolysis		GW697669
Otpt1_1489_AM21_024.ab1	1	Aminopeptidase, unknown function	622	25	8.07E-79	92.56 %	10	F:manganese ion binding; F:leucyl aminopeptidase activity; F:prolyl aminopeptidase activity; P:response to stress; F:aminopeptidase activity; C:chloroplast; P:proteolysis; F:zinc ion binding; P:arginine metabolic process; P:proline metabolic process	EC:3.4.11.1,	GW698088
Otpt1_1211_CH0101_080.ab1	1	Chloroplast nucleoid DNA binding; proteolysis	718	25	5.97E-51	69.24 %	4	F:pepsin A activity; P:proteolysis; F:aspartic-type endopeptidase activity; F:DNA binding	EC:3.4.23,	GW698015
Otpt1_1532_AM21_064.ab1	1	Signaling GTPase mediated ADP-ribosylation factor 1; GIP-binding; monomeric G	717	25	2.26E-66	99.80 %	8	F:GTP binding; C:Golgi apparatus; F:hydrolase activity; P:vesicle-mediated transport; P:intracellular		GW698105

		protein						protein transport; P:small GTPase mediated signal transduction; C:mitochondrion; F:transporter activity		
temperate.fasta.screen.Contig61	2	ADP-ribosylation factor-like protein	688	25	4.17E-83	93.84 %	3	C:endomembrane system; C:plasma membrane; C:vacuole	Otpt1_1406_CH0101_077.ab1; Otpt1_1471_AM21_004.ab1	GW698056 GW698080
Otpt1_0473_CZ1835_032.ab1	1	GDP dissociation inhibitor; protein transport	613	25	1.44E-107	94.28 %	3	F:Rab GDP-dissociation inhibitor activity; P:protein transport; P:regulation of GTPase activity		GW697730
Otpt1_0131_HFY328_086.ab1	1	RAB7 pruarras-related protein; signal transduction, protein transport and regulation of transcription;	535	25	1.81E-85	98.12 %	9	C:mitochondrion; P:small GTPase mediated signal transduction; P:regulation of transcription, DNA-dependent; F:ATP binding; F:transcription factor binding; P:protein transport; F:GTP binding; C:plasma membrane; C:transcription factor complex		GW697605
		Calcium related								
*Otpt1_1172_CH0101_034.ab1	1	Calmodulin – calcium bound calmodulin isoform 1; Voltage gated cation channel; Phosphatidylinositol signaling system ; response to stimulus	454	25	3.74E-14	100.00 %	1	F:calcium ion binding		GW698000
		Protein modifications – phosphorylation dephosphorylation								

Otpt1_0817_CZ1855_095.ab1	1	Casein kinase	486	14	8.71E-19	69.50 %	0		-	GW697857
Otpt1_0469_CZ1835_021.ab1	1	cysteine proteinase cp2; regulation of autophagy; endomembrane system;	677	25	7.04E-88	85.64 %	3	P:response to stress; P:proteolysis; F:cysteine-type endopeptidase activity	EC:3.4.22,	GW697728
Otpt1_1099_CH0101_083.ab1	1	EYES ABSENT protein; mTOR and ErbB signaling pathways, tyrosine phosphatase, metal dependant	576	25	9.58E-52	69.12 %	0			GW697961
		Nucleotides								
Otpt1_0409_CZ1835_065.ab1	1	Adenosine deaminase; purine biosynthesis	564	25	2.35E-68	71.40 %	4	P:nucleotide metabolic process; F:deaminase activity; F:hydrolase activity; P:purine ribonucleoside monophosphate biosynthetic process		GW697708
		Transport								
		Proteins transport								
Otpt1_0264_HFY330_072.ab1	1	ABC transporter protein 1-like - membrane	746	25	2.59E-100	84.56 %	2	F:ATP binding; F:ATPase activity	-	GW697653
Otpt1_0985_CH0101_016.ab1	1	SEC61 BETA (suppressors of secretion-defective 61 beta); protein transporter	464	25	1.31E-14	89.60 %	1	P:protein transport		GW697920
Otpt1_1149_CH0101_018.ab1	1	Developmental protein; protein transport with Snf7 domain	773	25	8.34E-81	90.56 %	1	P:protein transport		GW697989

Otpt1_1477_AM21_025.ab1	1	translocase inner mitochondrial membrane protein; subunit 23-2;	586	25	2.68E-49	74.40 %	1	C:plastid			GW698082
		Water transport									
Otpt1_0807_CZ1855_076.ab1	1	Aquaporin	221	0			0		-		GW697853
temperate.fasta.screen.Contig97	4	Transmembrane channel protein; aquaporin	655	25	2.91E-51	98.44 %	4	P:transport; C:plasma membrane; C:integral to membrane; F:transporter activity		Otpt1_0474_CZ1835_030.ab1; Otpt1_1081_CH0101_067.ab1; Otpt1_0922_CH0101_060.ab1; Otpt1_1386_CH0101_063.ab1	GW697731 GW697953 GW697898 GW698046
		Lipids transport									
Otpt1_0749_CZ1855_045.ab1	1	Acyl-CoA-binding protein (ACBP1); lipid transport; lead ion binding; mitochondria membrane	454	25	1.62E-17	89.00 %	2	F:acyl-CoA binding; P:transport			GW697836
temperate.fasta.screen.Contig18	1	Lipid-transfer protein precursor, facilitate transfer of fatty acid between membranes	427	25	6.86E-16	81.24 %	2	F:lipid binding; P:lipid transport		Otpt1_0030_HFY328_031.ab1	GW697569
temperate.fasta.screen.Contig31	1	Lipid-transfer protein precursor, facilitate transfer of fatty acid between membranes	416						-	Otpt1_0423_CZ1835_093.ab1	GW697713

temperate.fasta.screen.Contig2	1	Lipid-transfer protein precursor, facilitate transfer of fatty acid between membranes	411	23	1.24E-09	50.78 %	4	C:endomembrane system; P:lipid transport; F:lipid binding; P:transport	-	Otpt1_0692_CZ1855_096.ab1	GW697813
temperate.fasta.screen.Contig106	4	Non-specific lipid-transfer protein; endomembrane; lipid transfer; response to abscissic acid	646	25	4.81E-35	75.40 %	2	F:lipid binding; P:lipid transport		Otpt1_0565_CZ1855_015.ab1; Otpt1_0859_CH0101_004.ab1; Otpt1_0185_HFY330_041.ab1; Otpt1_0973_CH0101_084.ab1	GW697772 GW697872 GW697631 GW697915
*temperate.fasta.screen.Contig102	4	Non-specific lipid-transfer protein; endomembrane; lipid transfer; response to abscissic acid	623	25	2.31E-36	76.32 %	2	F:lipid binding; P:lipid transport		Otpt1_0498_CZ1835_044.ab1	GW697738
		Nucleotides transport									
Otpt1_0882_CH0101_018.ab1	1	ATP:ADP antiporter/ binding (AAC2) adenosine nucleotide translocator protein; mitochondira membrane; purine transport	409	25	1.61E-41	68.08 %	10	C:nucleolus; C:vacuole; F:binding; C:mitochondrial inner membrane; C:plasma membrane; C:cell wall; P:transport; C:integral to membrane; C:chloroplast; F:transporter activity			GW697879
		Hormone biosysntesis or hormone induced									

*temperate.fasta.screen.Contig40	2	ABA induced protein; pollen coat protein	531	22	1.69E-11	67.77 %	3	C:viral capsid; F:structural molecule activity; P:biological_process	-	Otpt1_0895_CH0101_046.ab1; Otpt1_1029_CH0101_041.ab1	GW697885 GW697935
*temperate.fasta.screen.Contig26	1	ABA induced protein; pollen coat protein	379	24	4.60E-12	67.96 %	3	C:viral capsid; F:structural molecule activity; P:biological_process	-	Otpt1_0488_CZ1835_039.ab1	GW697736
Otpt1_0662_CZ1855_073.ab1	1	Auxin-induced protein 6b	533	25	7.07E-34	77.44 %	2	P:auxin mediated signaling pathway; P:response to cold			GW697801
Otpt1_1422_CH0101_072.ab1	1	Auxin-induced protein	441	25	2.01E-31	78.16 %	0		-		GW698115
Otpt1_0920_CH0101_062.ab1	1	Auxin-induced protein; for transcription regulation	501	25	1.92E-40	71.60 %	2	P:regulation of transcription; F:protein binding			GW697897
*Otpt1_1075_CH0101_075.ab1	1	Auxin associated family protein; dormancy; unknown function	630	25	2.74E-40	76.40 %	1	P:auxin mediated signaling pathway	-		GW697951
Otpt1_1248_CH0101_086.ab1	1	Aminopeptidase like protein; auxin polar transport	377	17	3.19E-21	73.76 %	0				GW698024
Otpt1_0279_HFY330_094.ab1	1	Brassinosteroid biosynthetic protein, for unidimensional dell growth	713	25	8.60E-111	91.12 %	3	F:oxidoreductase activity; C:plasma membrane; F:FAD binding			GW697665
Otpt1_0894_CH0101_048.ab1	1	ACC oxydase; aminocyclopropane-1-carboxylate oxidase; ethylene response	711	25	2.51E-102	94.32 %	6	P:ethylene biosynthetic process; P:ripening; F:L-ascorbic acid binding; F:iron ion binding; F:1-aminocyclopropane-1-	EC:1.14.17.4,		GW697884

								carboxylate oxidase activity; P:oxidation reduction			
Otpt1_1483_AM21_032.ab1	1	Gibberellin regulated protein; gast-like gene product; for unidimensional dell growth; also regulated by brassinosteroid and abscissic acid	593	25	4.58E-28	82.24 %	0				GW698085
Otpt1_1148_CH0101_020.ab1	1	Gibberellin regulated protein; anti-microbial peptide, involved plant development	508	25	4.10E-30	84.12 %	1	P:response to gibberellin stimulus	-		GW697988
temperate.fasta.screen.Contig60	2	Gibberellin regulated protein; gip1-like protein	665	25	7.04E-53	77.20 %	0		-	Otpt1_1405_CH0101_079.ab1; Otpt1_1497_AM21_045.ab1	GW698055 GW698091
temperate.fasta.screen.Contig13	1	Gibberellin regulated protein; gip1-like protein	656	25	3.67E-54	77.88 %	0		-	Otpt1_0579_CZ1855_008.ab1	GW697777
temperate.fasta.screen.Contig50	2	Indole-3-acetic acid induced protein ARG-2 homolog; unknown function; response to stress	648	25	1.47E-23	61.36 %	1	P:response to stress		Otpt1_0217_HFY330_053.ab1; Otpt1_0685_CZ1855_087.ab1	GW697641 GW697810
Otpt1_0618_CZ1855_033.ab1	1	Lipoxygenase-3 LOX3, jasmonic acid and abscissic acid response, electron transport	562	25	6.37E-26	81.08 %	7	C:cytoplasm; P:oxylipin biosynthetic process; F:electron carrier activity; F:iron ion binding; P:oxidation reduction; F:lipoxygenase activity; P:electron transport	EC:1.13.11.12,		GW697791

temperate.fasta.screen.Contig79	3	Jasmonate induced protein	885	20	1.18E-71	60.75 %	1	C:mitochondrion		Otpt1_0876_CH0101_028.ab1; Otpt1_0879_CH0101_022.ab1; Otpt1_0771_CZ1855_063.ab1	GW697877 GW697878 GW697842
Otpt1_0701_CZ1855_084.ab1	1	Secondary metabolite Dihydroflavonol reductase; Flavonoid and lignin biosynthesis, secondary metabolite	701	25	5.68E-75	75.76 %	6	F:cinnamoyl-CoA reductase activity; P:lignin biosynthetic process; F:coenzyme binding; C:plastid; P:coumarin biosynthetic process; P:stilbene biosynthetic process			GW697819
Otpt1_0428_CZ1835_087.ab1		ENT-kaurenoic acid oxidase; diterpenoid biosynthesis, requires cytochrome P450	674	25	1.33E-38	68.40 %	2	F:binding; C:cell part			GW697715
Otpt1_0980_CH0101_007.ab1	1	Isoflavone reductase-like NADH-dependent oxidoreductase; negative regulation of transcription; regulation of nitrogen utilization; response to cadmium ions	675	25	2.37E-51	88.72 %	8	P:regulation of nitrogen utilization; C:cytoplasm; P:oxidation reduction; F:oxidoreductase activity; F:binding; F:transcription repressor activity; P:negative regulation of transcription; C:transcriptional repressor complex			GW697918
Otpt1_0133_HFY328_082.ab1	1	Oxidoreductase, 2OG-Fe(II) oxygenase family protein; leucoanthocyanidin dioxygenase	481	25	2.32E-32	85.16 %	2	F:oxidoreductase activity; F:iron ion binding	EC:1.21.3.1,		GW697607

Otpt1_0271_HFY330_089.ab1	1	SRG1 (senescence-related gene 1); oxidoreductase, a 2og-Fe(II) oxygenase, flavonoid biosynthetis,	682	25	7.33E-24	87.36 %	2	P:flavonoid biosynthetic process; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors	-	GW697659
Otpt1_0720_CZ1855_006.ab1	1	Oxydoreductase; stilbene and flavonoid metabolism	588	25	2.63E-44	71.68 %	2	P:flavonoid biosynthetic process; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors	-	GW697827
Otpt1_0413_CZ1835_078.ab1	1	Oxydoreductase; stilbene and flavonoid metabolism	546	25	5.55E-48	68.64 %	2	P:flavonoid biosynthetic process; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors	-	GW697709
Otpt1_1478_AM21_023.ab1	1	Senescence-related gene 1 (srg1); incorporation or reduction of molecular 2-oxoglutarate; stilbene and flavonoid metabolism	657	25	1.56E-72	74.24 %	2	P:flavonoid biosynthetic process; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors		GW698083
		Response to biotic or abiotic stimulus								

temperate.fasta.screen.Contig91	3	ADR6 aluminum induced; embryonic abundant protein precursor-like protein; with BURP domain	741	25	3.13E-106	58.44 %	0		-	Otpt1_0753_CZ1855_041.ab1; Otpt1_1010_CH0101_030.ab1; Otpt1_0526_CZ1835_077.ab1	GW697837 GW697927 GW697752
temperate.fasta.screen.Contig95	3	ADR6 aluminum induced; embryonic abundant protein precursor-like protein; with BURP domain	712	25	6.86E-84	57.40 %	0		-	Otpt1_0182_HFY330_047.ab1; Otpt1_1028_CH0101_043.ab1; Otpt1_1441_CH0101_083.ab1	GW697628 GW697934 GW698063
temperate.fasta.screen.Contig52	2	ADR6 aluminum induced; embryonic abundant protein precursor-like protein; with BURP domain	712	25	5.07E-87	56.20 %	0		-	Otpt1_1544_AM21_077.ab1; Otpt1_0563_CZ1835_084.ab1	GW698110 GW697771
*Otpt1_0708_CZ1855_005.ab1	1	AVR99 cf-9 rapidly elicited protein 20; calcium binding	527	25	5.74E-41	81.32 %	3	P:protein folding; C:cytosol; F:calcium ion binding			GW697825
temperate.fasta.screen.Contig111	5	Cold-induced PsAD2-like protein	664	3	8.11E-25	70.33 %	0		-	Otpt1_0837_CZ1855_084.ab1; Otpt1_1206_CH0101_073.ab1; Otpt1_0604_CZ1855_026.ab1; Otpt1_0941_CH0101_080.ab1; Otpt1_0679_CZ1855_091.ab1	GW697862 GW698013 GW697786 GW697907 GW697808
temperate.fasta.screen.Contig22	1	Cold-induced PsAD2-like protein	522	2	1.25E-19	72.00 %	0		-	Otpt1_1194_CH0101_058.ab1	GW698007

temperate.fasta.screen.Contig3	1	Cold induced PsAD2-like	476	3	4.26E-26	66.67 %	0		-	Otpt1_1262_CH0101_014.ab1	GW698030
Otpt1_0137_HFY330_009.ab1	1	Senescence-associated protein; cold regulated	535	10	3.44E-52	84.70 %	0		-		GW697609
Otpt1_0864_CH0101_025.ab1	1	Dehydrin, very weak similarity to ERD10 and ERD14 COR protein	635	1	9.96E-06	88.00 %	2	P:response to water; P:response to stress			GW697874
Otpt1_0238_HFY330_052.ab1	1	Drought-stress protein	463	0			0		-		GW697646
*temperate.fasta.screen.Contig64	2	Germin-like protein; Mn ion binding, nutrient reservoir	469	25	3.29E-26	78.00 %	7	F:nutrient reservoir activity; F:manganese ion binding; C:cytoplasmic membrane-bounded vesicle; F:superoxide dismutase activity; P:oxidation reduction; C:apoplast; P:superoxide metabolic process		Otpt1_0337_CZ1835_018.ab1; Otpt1_0434_CZ1835_092.ab1	GW697688 GW697719
*Otpt1_1500_AM21_039.ab1	1	Germin-like protein; Mn ion binding, nutrient reservoir	394	25	3.42E-47	78.12 %	2	C:extracellular region; F:manganese ion binding			GW698093
temperate.fasta.screen.Contig87	3	Heat- and acid-stable phosphoprotein 28 kda (HASP)	765	25	6.14E-44	78.56 %	2	C:mitochondrion; P:cell proliferation		Otpt1_1123_CH0101_012.ab1; Otpt1_1203_CH0101_079.ab1; Otpt1_0592_CZ1855_029.ab1	GW697975 GW698010 GW697779
Otpt1_0896_CH0101_044.ab1	1	Heat shock protein 90; ATP binding	559	25	4.00E-81	97.44 %	6	P:protein folding; F:unfolded protein binding; C:cytoplasm; F:ATP binding; P:response to stress; C:plasma membrane	-		GW697886

Otpt1_0549_CZ1835_087.ab1	1	Heat-shock protein; endomembrane system	559	9	3.00E-16	77.89 %	0		-		GW697768
Otpt1_0080_HFY328_062.ab1	1	Heat shock protein DnaJ	705	12	2.00E-11	74.08 %	1	F:heat shock protein binding	-		GW697589
Otpt1_0548_CZ1835_089.ab1	1	Light-inducible protein ATLS1; macrophage migration inhibitory factor family protein (MIF); response to other organisms	674	25	5.57E-53	91.48 %	3	P:inflammatory response; C:membrane; P:response to other organism			GW697767
*Otpt1_0104_HFY328_072.ab1	1	Pathogenesis related PR10-1 protein	708	25	1.21E-56	80.88 %	3	P:defense response; P:response to biotic stimulus; C:cytoplasm			GW697596
*temperate.fasta.screen.Contig54	2	Pathogenesis related protein; ABR18	721	25	1.75E-66	84.32 %	3	P:defense response; P:response to biotic stimulus; C:cytoplasm		Otpt1_1367_CH0101_041.ab1; Otpt1_0569_CZ1855_009.ab1	GW698039 GW697775
Otpt1_1212_CH0101_078.ab1	1	Pathogen related	504	2	4.42E-16	61.50 %	0		-		GW698016
temperate.fasta.screen.Contig34	1	Pathogen related	747	25	2.21E-99	73.52 %	2	P:response to biotic stimulus; P:defense response	-	Otpt1_0519_CZ1835_058.ab1	GW697749
temperate.fasta.screen.Contig6	1	Pathogen related	700	25	3.42E-96	75.72 %	2	P:response to biotic stimulus; P:defense response	-	Otpt1_0669_CZ1855_080.ab1	GW697804
temperate.fasta.screen.Contig101	4	Ripening related protein; bet v 1; major latex-like protein	743	25	2.76E-78	66.08 %	3	P:defense response; P:response to biotic stimulus; C:chloroplast		Otpt1_1473_AM21_031.ab1; Otpt1_1042_CH0101_038.ab1; Otpt1_1380_CH0101	GW698081 GW697941 GW698045 GW697795

										_042.ab1; Otpt1_0634_CZ1855_061.ab1	
temperate.fasta.screen.Contig81	3	Ripening related protein; bet v 1; major latex-like protein	734	25	7.32E-84	67.12 %	3	P:defense response; P:response to biotic stimulus; C:chloroplast		Otpt1_1465_AM21_014.ab1; Otpt1_0017_HFY328_005.ab1; Otpt1_1169_CH0101_036.ab1	GW698077 GW697566 GW697999
temperate.fasta.screen.Contig93	3	Ripening related protein; bet v 1; major latex-like protein	717	25	1.41E-76	65.00 %	3	P:defense response; P:response to biotic stimulus; C:chloroplast		Otpt1_0323_CZ1835_021.ab1; Otpt1_0468_CZ1835_023.ab1; Otpt1_1254_CH0101_009.ab1	GW697681 GW697727 GW698027
temperate.fasta.screen.Contig48	2	Ripening related protein; bet v 1	642	25	1.57E-62	64.32 %	2	P:response to biotic stimulus; P:defense response		Otpt1_1368_CH0101_039.ab1; Otpt1_0193_HFY330_048.ab1	GW698040 GW697633
temperate.fasta.screen.Contig32	1	Ripening related protein; bet v 1	679	25	5.37E-83	66.84 %	3	P:defense response; P:response to biotic stimulus; C:chloroplast		Otpt1_0502_CZ1835_036.ab1	GW697741
temperate.fasta.screen.Contig17	1	Ripening related protein; bet v 1; major latex-like protein	651	25	1.09E-70	65.68 %	3	P:defense response; P:response to biotic stimulus; C:chloroplast		Otpt1_0534_CZ1835_080.ab1	GW697758
temperate.fasta.screen.Contig8	1	Ripening related protein; bet v 1; major latex-like protein	610	25	4.15E-83	66.60 %	3	P:defense response; P:response to biotic stimulus; C:chloroplast		Otpt1_0566_CZ1855_013.ab1	GW697773
Otpt1_1511_AM21_044.ab1	1	Wound-response protein	524	6	5.08E-13	60.50 %	1	P:biological_process	-		GW698096
temperate.fasta.screen.Contig80	3	Vegetative storage protein; acid phosphatase	703	25	4.77E-73	70.92 %	3	F:acid phosphatase activity; P:riboflavin metabolic process;	EC:3.1.3.2,	Otpt1_1068_CH0101_058.ab1; Otpt1_0872_CH0101	GW697947 GW697876 GW697691

								P:hexachlorocyclohexane metabolic process		_030.ab1; Otpt1_0353_CZ1835_044.ab1	
Otpt1_0031_HFY328_029.ab1	1	Vegetative storage protein; acid phosphatase	709	25	4.70E-85	63.00 %	1	C:cellulose and pectin-containing cell wall	EC:3.1.3.2,		GW697570
Otpt1_1461_AM21_005.ab1	1	Vegetative storage protein; acid phosphatase	636	25	8.65E-74	64.12 %	3	C:cellulose and pectin-containing cell wall; C:cytoplasmic part; C:intracellular membrane-bounded organelle	EC:3.1.3.2,		GW698073
Otpt1_0672_CZ1855_074.ab1	1	Tetratricopeptide repeat-containing protein;	185	25	3.35E-18	72.68 %	4	F:molecular_function; F:binding; P:biological_process; C:cellular_component			-
Otpt1_0851_CH0101_001.ab1	1	Trypsin protein inhibitor 2; defense response	656	25	9.97E-60	62.00 %	2	C:extracellular region; F:serine-type endopeptidase inhibitor activity			GW697866
*Otpt1_0075_HFY328_053.ab1	1	Trypsin protein inhibitor 3	702	25	7.62E-80	53.32 %	2	F:peptidase activity; F:endopeptidase inhibitor activity			GW697587
		Lipids									
Otpt1_0382_CZ1835_056.ab1	1	GDSL-motif lipase/hydrolase family protein, carboxylesterase; endomembrane	559	25	1.43E-46	72.04 %	2	F:lipase activity; P:lipid catabolic process	EC:3.1.1, EC:3.1.1.1,		GW697701
Otpt1_0100_HFY328_080.ab1	1	GDSL-motif lipase/hydrolase family protein, carboxylesterase; endomembrane	641	25	2.25E-37	72.36 %	0		-		GW697595

Otpt1_1138_CH0101_023.ab1	1	ATLIP1; lysosomal acid lipase; galactolipase/ hydrolase/ phospholipase/ triacylglycerol lipase	437	25	7.22E-26	70.08 %	0				GW697982
temperate.fasta.screen.Contig70	2	SEC14 cytosolic factor family protein; phosphoglyceride transfer family protein	525	25	1.32E-24	73.40 %	3	C:plastid; P:transport; F:transporter activity		Otpt1_0397_CZ1835_079.ab1; Otpt1_0504_CZ1835_063.ab1	GW697705 GW697743
Otpt1_0056_HFY328_037.ab1	1	lipid transfer protein family protein; LTP; protease inhibitor; seed storage	455	25	4.69E-41	82.32 %	2	P:lipid transport; C:cytoplasmic membrane-bounded vesicle	-		GW697577
		Pyruvate - fatty acids									
*temperate.fasta.screen.Contig41	2	accD acetyl CoA carboxylase beta subunit	521	25	4.41E-86	91.36 %	7	C:acetyl-CoA carboxylase complex; P:fatty acid biosynthetic process; F:transferase activity; C:chloroplast; F:zinc ion binding; F:acetyl-CoA carboxylase activity; P:pyruvate metabolic process	EC:6.4.1.2,	Otpt1_0160_HFY330_025.ab1; Otpt1_0483_CZ1835_018.ab1	GW697619 GW697735
		Ribosomes and translation									
		Ribosome biogenesis and assembly - organelles									

Otpt1_0961_CH0101_083.ab1	1	50s ribosomal protein L12 chloroplast precursor	691	25	2.96E-36	80.68 %	9	C:ribosome; F:structural constituent of ribosome; F:RNA binding; C:stromule; C:nucleoid; C:chloroplast; C:thylakoid; P:translation; P:ribosome biogenesis and assembly	EC:3.6.5.3,	GW697912
*Otpt1_1376_CH0101_033.ab1	1	23S ribosomal RNA; orf9; chloroplast genome	488	25	3.75E-46	76.00 %	1	C:chloroplast		GW698042
		Ribosome biogenesis and assembly - cell								
Otpt1_1141_CH0101_017.ab1	1	40s ribosomal protein s11	682	25	2.40E-75	95.68 %	6	F:structural constituent of ribosome; C:cell wall; C:cytosolic small ribosomal subunit; P:translation; F:rRNA binding; P:ribosome biogenesis and assembly	EC:3.6.5.3,	GW697984
Otpt1_0706_CZ1855_009.ab1	1	40S ribosomal protein s13	716	25	5.68E-78	96.04 %	4	F:structural constituent of ribosome; C:cytosolic small ribosomal subunit; P:translation; P:ribosome biogenesis and assembly	EC:3.6.5.3,	GW697823
Otpt1_1072_CH0101_050.ab1	1	40S ribosomal protein s3a	698	25	9.25E-86	94.84 %	4	C:ribosome; F:structural constituent of ribosome; P:translation; P:ribosome biogenesis and assembly	EC:3.6.5.3,	GW697950
Otpt1_0610_CZ1855_018.ab1	1	40S ribosomal protein s3a	642	25	1.04E-82	94.84 %	4	F:structural constituent of ribosome; C:cytosolic small ribosomal subunit; P:translation; P:ribosome biogenesis and assembly	EC:3.6.5.3,	GW697788
Otpt1_0123_HFY328_096.ab1	1	40S ribosomal protein s6	595	25	1.79E-96	96.48 %	6	C:nucleolus; C:cytosolic ribosome; F:structural constituent of ribosome; C:plasma membrane;	EC:3.6.5.3,	GW697602

								P:translation; P:ribosome biogenesis and assembly			
temperate.fasta.screen.Contig23	1	40s ribosomal protein s10-like	724	25	2.56E-49	89.28 %	2	C:cytosolic small ribosomal subunit; P:translation	EC:3.6.5.3,	Otpt1_0834_CZ1855_090.ab1	GW697861
Otpt1_0033_HFY328_025.ab1	1	40s ribosomal protein s15a	631	25	3.11E-68	98.32 %	6	F:structural constituent of ribosome; C:plasma membrane; C:cell wall; C:cytosolic small ribosomal subunit; P:translation; P:ribosome biogenesis and assembly	EC:3.6.5.3,		GW697571
Otpt1_1398_CH0101_060.ab1	1	40s ribosomal protein s9	717	25	2.83E-93	95.92 %	5	F:structural constituent of ribosome; C:cytosolic small ribosomal subunit; P:translation; F:rRNA binding; P:ribosome biogenesis and assembly	EC:3.6.5.3,		GW698051
*Otpt1_0221_HFY330_051.ab1	1	40s ribosomal protein s23	636	25	2.06E-75	97.04 %	5	F:structural constituent of ribosome; C:cytosolic small ribosomal subunit; P:translation; C:mitochondrion; P:ribosome biogenesis and assembly	EC:3.6.5.3,		GW697642
Otpt1_0388_CZ1835_054.ab1	1	60S ribosomal protein l1	647	25	7.77E-110	92.32 %	4	F:structural constituent of ribosome; C:cytosolic large ribosomal subunit; P:translation; P:ribosome biogenesis and assembly	-		GW697702
Otpt1_0897_CH0101_042.ab1	1	60S ribosomal protein l10a	735	25	8.68E-101	94.48 %	6	F:structural constituent of ribosome; P:post-embryonic organ development; P:adaxial/abaxial pattern formation; C:cytosolic large ribosomal subunit; P:translation; P:ribosome	-		GW697887

								biogenesis and assembly			
temperate.fasta.screen.Contig107	4	60S ribosomal protein l11 or RPL16	706	25	5.70E-83	97.32 %	5	F:structural constituent of ribosome; C:cytosolic large ribosomal subunit; P:translation; F:rRNA binding; P:ribosome biogenesis and assembly	EC:3.6.5.3,	Otpt1_0552_CZ1835_085.ab1; Otpt1_1409_CH0101_075.ab1; Otpt1_0721_CZ1855_004.ab1; Otpt1_0366_CZ1835_055.ab1	GW697769 GW698057 GW697828 GW697698
Otpt1_0628_CZ1855_040.ab1	1	60S ribosomal protein l13 or bbc1 protein	632	25	2.21E-98	93.60 %	6	C:cytosolic ribosome; F:structural constituent of ribosome; C:plasma membrane; C:cell wall; P:translation; P:ribosome biogenesis and assembly	EC:3.6.5.3,		GW697793
temperate.fasta.screen.Contig59	2	60S ribosomal protein l14 or glycoprotein-like ribosomal	591	25	3.04E-56	93.20 %	5	F:structural constituent of ribosome; P:ribosome biogenesis and assembly; C:cytosolic large ribosomal subunit; P:translation; C:endoplasmic reticulum	-	Otpt1_0343_CZ1835_039.ab1; Otpt1_0280_HFY330_092.ab1	GW697689 GW697666
Otpt1_1360_CH0101_047.ab1	1	60S ribosomal protein l17 domain; L14p/L23e; wound inducible	575	25	3.81E-69	97.56 %	5	F:structural constituent of ribosome; C:cytosolic large ribosomal subunit; P:translation; C:mitochondrion; P:ribosome biogenesis and assembly	EC:3.6.5.3,		GW698037
Otpt1_0266_HFY330_068.ab1	1	60S ribosomal protein l18	346	25	1.39E-24	97.20 %	6	C:vacuole; F:structural constituent of ribosome; C:plasma membrane; C:cytosolic large ribosomal subunit; P:translation; P:ribosome biogenesis and assembly	EC:3.6.5.3,		GW697655

Otpt1_0132_HFY328_084.ab1	1	60S ribosomal protein l27 homolog	526	25	4.42E-34	88.48 %	5	F:structural constituent of ribosome; P:ribosome biogenesis and assembly; C:cytosolic large ribosomal subunit; P:translation; C:mitochondrion	EC:3.6.5.3,	GW697606
Otpt1_0183_HFY330_045.ab1	1	60S ribosomal protein l30	573	25	1.96E-57	94.40 %	4	F:structural constituent of ribosome; C:cytosolic large ribosomal subunit; P:translation; P:ribosome biogenesis and assembly	EC:3.6.5.3,	GW697629
Otpt1_0363_CZ1835_061.ab1	1	60S ribosomal protein L35a	591	25	1.91E-58	94.80 %	4	F:structural constituent of ribosome; P:ribosome biogenesis and assembly; C:cytosolic large ribosomal subunit; P:translation	EC:3.6.5.3,	GW697695
Otpt1_0454_CZ1835_016.ab1	1	60S ribosomal protein l38	469	25	1.29E-30	96.44 %	4	F:structural constituent of ribosome; P:ribosome biogenesis and assembly; C:cytosolic large ribosomal subunit; P:translation	EC:3.6.5.3,	GW697722
Otpt1_1134_CH0101_025.ab1	1	60S ribosomal protein l7-like protein	779	25	1.31E-73	69.48 %	4	F:structural constituent of ribosome; C:cytosolic large ribosomal subunit; P:translation; P:ribosome biogenesis and assembly	-	GW697981
Otpt1_0730_CZ1855_025.ab1	1	ubiquitin extension protein; ribosomal; S27A; response to dehydration	623	25	5.10E-68	98.40 %	5	C:ribosome; P:protein modification process; F:structural constituent of ribosome; P:translation; P:ribosome biogenesis and assembly	EC:3.6.5.3,	GW697829
Otpt1_1288_CH0101_032.ab1	1	ubiquitin fusion protein; fused to ribosomal protein L40	541	25	5.68E-66	95.56 %	6	P:protein modification process; F:structural constituent of ribosome; C:cytosolic large ribosomal subunit; C:nucleus;	EC:3.6.5.3,	GW698033

								P:translational elongation; P:ribosome biogenesis and assembly		
Otpt1_1139_CH0101_021.ab1	1	ubiquitin fusion protein	648	25	1.07E-66	97.36 %	5	P:protein modification process; F:structural constituent of ribosome; C:cytosolic large ribosomal subunit; P:translational elongation; P:ribosome biogenesis and assembly	EC:3.6.5.3,	GW697983
		Other translation proteins								
Otpt1_0933_CH0101_075.ab1	1	Eukaryotic translation initiation factor 4e	356	25	8.09E-17	94.52 %	4	F:RNA binding; F:translation initiation factor activity; C:ribosome; P:regulation of translational initiation		GW697902
Otpt1_1488_AM21_026.ab1	1	Translation release factor	477	25	1.39E-69	81.60 %	5	F:translation release factor activity; C:ribosome; ; P:regulation of translational termination; C:translation release factor complex		GW698087
Otpt1_0528_CZ1835_073.ab1	1	Elongation factor Tu; chloroplast precursor	440	25	1.50E-23	94.00 %	6	F:translation elongation factor activity; F:GTP binding; C:chloroplast; F:GTPase activity; C:ribosome; P:regulation of translational elongation	EC:3.6.5.1, EC:3.6.5.2, EC:3.6.5.3, EC:3.6.5.4,	GW697754
Otpt1_0496_CZ1835_048.ab1	1	EMBRYO DEFECTIVE 1129; structural constituent of ribosome	553	0			0		-	GW697737
		RNA associated								

		proteins								
Otpt1_1030_CH0101_039.ab1	1	RNA recognition motif-containing protein; similar to DEAD-box helicase	720	25	5.29E-39	77.20 %	1	C:mitochondrion		GW697936
Otpt01_0002_CZ1321_041.ab1	1	RNA-binding protein; chloroplast; polyU binding	689	25	2.23E-52	69.28 %	5	P:innate immune response; P:response to cold; C:stromule; F:poly(U) binding; C:chloroplast		GW698111
Otpt1_0815_CZ1855_066.ab1	1	RAS-gtpase-activating protein sh3-domain binding protein; RRM domain for RNA binding; traffic between cytoplasm and nucleus	552	25	1.24E-50	61.96 %	1	P:nucleocytoplasmic transport	-	GW697856
Otpt1_1481_AM21_019.ab1	1	RNA-binding s4 domain-containing protein	745	25	5.43E-82	85.44 %	3	C:mitochondrion; F:rRNA binding; C:cytosolic small ribosomal subunit	-	GW698084
Otpt1_0121_HFY328_083.ab1	1	RNS2 (ribonuclease 2) endoribonuclease; intracellular space	628	25	3.05E-92	71.60 %	4	C:cytoplasmic membrane-bounded vesicle; F:RNA binding; F:endoribonuclease activity; P:regulation of RNA metabolic process		GW697601
Otpt1_0245_HFY330_075.ab1	1	Small nucleolar ribonucleotide complex ; nucleus	648	25	6.33E-51	96.84 %	1	C:small nucleolar ribonucleoprotein complex		GW697648
Otpt1_1151_CH0101_045.ab1	1	Small nucleolar ribonucleoprotein F; nucleus	665	25	2.97E-35	95.44 %	1	C:nucleolus	-	GW697990
Otpt1_0529_CZ1835_071.ab1	1	Small nucleolar ribonucleotide complex F; nucleus	625	25	8.90E-36	85.72 %	7	C:nucleolus; P:RNA splicing; C:spliceosome; F:RNA binding; F:protein binding;		GW697990

								C:small nuclear ribonucleoprotein complex; P:mRNA processing		
		DNA associated proteins								
Otpt1_1160_CH0101_046.ab1	1	Sister-chromatid cohesion protein; cell division and chromosome partitioning	698	12	5.34E-25	70.58 %	0		-	GW697996
Otpt1_0542_CZ1835_068.ab1	1	TINY-like protein; with AP2-domain ; DNA binding; ERF023 ethylene response factor	706	25	5.96E-40	74.76 %	3	F:transcription factor activity; P:regulation of transcription, DNA-dependent; C:transcription factor complex		GW697762
Otpt1_0758_CZ1855_037.ab1	1	EREBP4 ethylene-responsive element-binding protein; transcription factor AP2/EREBP family	623	25	1.54E-56	86.24 %	0		-	GW697838
Otpt1_0088_HFY328_050.ab1	1	PTAC2 (plastid transcriptionally active2); transcription factor for chloroplast chromosome	634	13	4.56E-51	70.31 %	2	C:nucleoid; C:plastid		GW697591
Otpt1_1001_CH0101_023.ab1	1	Homeodomain transcription factor class III HD-Zip protein; PHB family	596	25	4.36E-18	88.00 %	6	F:sequence-specific DNA binding; F:protein dimerization activity; F:transcription factor activity; C:mitochondrion; P:regulation of transcription, DNA-dependent; C:transcription factor complex		GW697924

temperate.fasta.screen.Contig43	2	Double-stranded DNA-binding family protein	687	25	1.23E-34	82.12 %	1	F:double-stranded DNA binding		Otpt1_0773_CZ1855_059.ab1; Otpt1_1167_CH0101_038.ab1	GW697844 GW697998
Otpt1_0861_CH0101_031.ab1	1	DVL10/rotundifolia - weak similarity	489	0			0		-		GW697873
Otpt1_0209_HFY330_063.ab1	1	VCS (VARICOSE) nucleotide binding; with WD40 decapping of mRNA domain	635	25	2.21E-13	58.16 %	1	C:plastid			GW697638
Otpt1_0037_HFY328_019.ab1	1	Zinc finger (B-box type) family protein; transcription factor	562	25	4.55E-40	69.64 %	3	P:regulation of transcription; F:transcription factor activity; C:transcription factor complex			GW697573
Otpt1_0115_HFY328_095.ab1	1	Zinc finger (C2H2 type) family protein; transcription factor	576	25	9.28E-47	82.64 %	5	P:regulation of transcription; F:transcription factor activity; F:zinc ion binding; C:mitochondrion; C:transcription factor complex			GW697599
		Histones									
Otpt1_1130_CH0101_031.ab1	1	Histone deacetylase HD2a	702	25	3.26E-22	75.44 %	3	F:metal ion binding; C:intracellular; P:transcription			GW697978
*Otpt1_0616_CZ1855_037.ab1	1	Weakely similar to histone deacetylase Otpt1_1130	533	0			0		-		GW697790
Otpt1_1463_AM21_001.ab1	1	Histone h1	602	25	6.24E-28	87.12 %	6	C:nucleosome; F:DNA binding; P:nucleosome assembly; P:response to stress; C:nucleus;			GW698075

								C:mitochondrion			
temperate.fasta.screen.Contig85	3	Histone h2a	829	25	2.14E-40	98.68 %	5	C:nucleosome; F:DNA binding; P:nucleosome assembly; C:nucleus; C:plastid		Otpt1_0505_CZ1835_061.ab1; Otpt1_0506_CZ1835_059.ab1; Otpt1_0307_CZ1835_010.ab1	GW697744 GW697745 GW697676
temperate.fasta.screen.Contig65	2	Histone h2a	736	25	1.19E-41	97.48 %	6	C:nucleosome; F:DNA binding; P:nucleosome assembly; C:nucleus; C:plastid; C:mitochondrion		Otpt1_0842_CH0101_013.ab1; Otpt1_1427_CH0101_095.ab1	GW697864 GW698062
temperate.fasta.screen.Contig29	1	Histone h2a	656	25	3.94E-40	98.68 %	5	C:nucleosome; F:DNA binding; P:nucleosome assembly; C:nucleus; C:plastid		Otpt1_0196_HFY330_044.ab1	GW697635
Otpt1_0163_HFY330_023.ab1	1	Histone h2b	753	25	6.50E-43	95.80 %	4	C:nucleosome; F:DNA binding; P:nucleosome assembly; C:nucleus			GW697620
temperate.fasta.screen.Contig71	2	Histone h2b	696	25	5.74E-46	99.84 %	4	C:nucleosome; F:DNA binding; P:nucleosome assembly; C:nucleus	-	Otpt1_1050_CH0101_059.ab1; Otpt1_0673_CZ1855_072.ab1	GW697943 GW697805
Otpt1_1142_CH0101_032.ab1	1	Histone h2b	432	25	4.86E-22	100.00 %	4	C:nucleosome; F:DNA binding; P:nucleosome assembly; C:nucleus	-		GW697985
Otpt1_1154_CH0101_039.ab1	1	Histone h3	719	25	2.41E-68	97.80 %	6	C:nucleosome; F:DNA binding; P:nucleosome assembly; C:nucleus; C:plastid; C:mitochondrion			GW697993
Otpt1_0733_CZ1855_019.ab1	1	Histone h4	530	25	1.44E-39	99.64 %	4	C:nucleosome; F:DNA binding; P:nucleosome assembly; C:nucleus			GW697831
Otpt1_0965_CH0101_092	1	Histone h4	526	25	7.44E-36	98.88 %	4	C:nucleosome; F:DNA binding; P:nucleosome			GW697913

.ab1								assembly; C:nucleus		
		ROS scavenging and signaling								
Otpt1_0500_CZ1835_040.ab1	1	Lactoylglutathione lyase family protein glyoxalase 1 family protein	726	25	1.04E-74	76.88 %	1	P:metabolic process		GW697739
*Otpt1_0294_CZ1835_003.ab1	1	Glutathione peroxidase; Phospholipid hydroperoxide glutathione peroxidase 1; chloroplast	692	25	4.20E-91	89.48 %	8	C:chloroplast envelope; P:response to oxidative stress; F:phospholipid-hydroperoxide glutathione peroxidase activity; P:oxidation reduction; C:stromule; F:glutathione peroxidase activity; P:glutathione metabolic process; P:peroxidase reaction	EC:1.11.1.12, EC:1.11.1.9,	GW697673
*Otpt1_1358_CH0101_018.ab1	1	glutathione peroxidase; phospholipid hydroperoxide glutathione peroxidase	724	25	2.99E-82	88.40 %	7	F:phospholipid-hydroperoxide glutathione peroxidase activity; F:glutathione peroxidase activity; C:cytoplasm; P:oxidation reduction; P:response to oxidative stress; P:glutathione metabolic process; P:peroxidase reaction	EC:1.11.1.9,	GW698036
Otpt1_0086_HFY328_054.ab1	1	Similar to GRF6 (G-BOX REGULATING FACTOR 6) in brassinosteroid mediated signaling	687	25	6.86E-94	91.20 %	5	P:response to cadmium ion; C:cell wall; C:chloroplast; C:plasma membrane; F:protein domain specific binding		GW697590

Otpt1_1208_CH0101_069.ab1	1	Fe-superoxide dismutase 2 precursor	556	25	8.24E-55	81.08 %	6	P:oxidation reduction; P:superoxide metabolic process; F:iron ion binding; F:superoxide dismutase activity; C:nucleoid; C:chloroplast	EC:1.15.1.1,	GW698014
Otpt1_1514_AM21_040.ab1	1	Fe-superoxide dismutase 2 precursor	369	25	1.35E-35	84.92 %	10	P:oxidation reduction; P:superoxide metabolic process; F:iron ion binding; F:superoxide dismutase activity; C:nucleoid; C:chloroplast; C:mitochondrion; F:oxidoreductase activity; F:metal ion binding; C:plasma membrane	1.35E-35	GW698097
*Otpt1_0016_HFY328_007.ab1	1	Peroxiredoxin q; thylakoid; thioredoxin reductase; antioxydant and protein folding	713	25	5.99E-88	82.52 %	4	F:peroxiredoxin activity; P:oxidation reduction; C:plastoglobule; F:protein binding	EC:1.11.1.15,	GW697565
Otpt1_0828_CZ1855_081.ab1	1	Glutathione s-transferase gst	368	25	2.92E-19	71.84 %	3	F:glutathione transferase activity; P:glutathione metabolic process; P:glutathione conjugation reaction	EC:2.5.1.18,	GW697859
Otpt1_0530_CZ1835_069.ab1	1	Glutathione s-transferase	700	25	2.25E-47	78.52 %	10	F:glutathione transferase activity; P:response to cadmium ion; C:plasma membrane; C:cell wall; C:stromule; C:chloroplast; C:thylakoid; C:apoplast; P:glutathione metabolic process; P:glutathione conjugation reaction	EC:2.5.1.18,	GW697756

Otpt1_1400_CH0101_056.ab1	1	Glutathione S-transferaseor Thioredoxin (TRX) superfamily;	497	25	1.17E-79	81.84 %	5	P:response to cadmium ion; F:glutathione transferase activity; C:plastid; P:glutathione metabolic process; P:glutathione conjugation reaction	EC:2.5.1.18,		GW698053
temperate.fasta.screen.Contig46	2	Glutaredoxin; arsenate reductase; Protein folding	703	25	2.57E-35	79.08 %	4	P:cell redox homeostasis; F:electron carrier activity; F:protein disulfide oxidoreductase activity; P:electron transport		Otpt1_0998_CH0101_027.ab1; Otpt1_1237_CH0101_083.ab1	GW697923 GW698019
Otpt1_1426_CH0101_066.ab1	1	Peroxidase; endomembrane	346	25	5.87E-15	87.16 %	10	F:heme binding; P:oxidation reduction; F:iron ion binding; P:response to nematode; F:peroxidase activity; F:electron carrier activity; F:calcium ion binding; P:response to oxidative stress; P:peroxidase reaction; P:electron transport	EC:1.11.1.7,		GW698061
Otpt1_1419_CH0101_078.ab1	1	Very weak similarity to a peroxydase	368	0			0		-		GW698060
		Unclassified									
Otpt1_0748_CZ1855_047.ab1	1	Dienelactone hydrolase family protein	752	25	7.83E-97	79.56 %	3	F:hydrolase activity; C:nucleus; C:cytoplasm			GW697835
Otpt1_0431_CZ1835_081.ab1	1	Flowering locus t-like protein (FT), protein and phosphatidylethanolamine binding, photoperiod and flowering, cytoplasm and	664	25	9.47E-66	82.12 %	0		-		GW697717

		nucleus								
Otpt1_1069_CH0101_056.ab1	1	Phosphate responsive 1 family; located in cell wall	456	25	3.03E-32	86.56 %	2	C:cytoplasmic membrane-bounded vesicle; C:cellulose and pectin-containing cell wall	-	GW697948
Otpt1_0373_CZ1835_062.ab1	1	CP5; membrane related protein	599	25	2.92E-70	80.00 %	1	P:biological_process	-	GW697699
Otpt1_0517_CZ1835_060.ab1	1	SOUL-like protein; Heme binding protein;	662	25	3.95E-48	59.68 %	1	P:biological_process	-	GW697748
Otpt1_0924_CH0101_056.ab1	1	2-oxoglutarate-dependent dioxygenase	547	25	6.35E-44	70.08 %	1	F:oxidoreductase activity		GW697900
Otpt1_1046_CH0101_034.ab1	1	Ubiquinone biosynthesis; similar to Coenzyme Q9; electron transport; mitochondrion	771	25	2.36E-67	65.04 %	1	C:mitochondrion	-	GW697942
Otpt1_0594_CZ1855_025.ab1	1	Flowering promoting factor-like 1; plastid	676	25	4.18E-48	78.76 %	2	C:mitochondrion; C:plastid	-	GW697780
Otpt1_1176_CH0101_059.ab1	1	Ankyrin repeat family	383	11	1.93E-10	61.55 %	3	F:molecular_function; P:biological_process; F:protein binding	-	GW698003
Otpt1_1525_AM21_059.ab1	1	Zinc binding; unknown function	542	25	1.32E-30	77.72 %	6	F:metal ion binding; F:iron-sulfur cluster binding; C:organelle membrane; C:endoplasmic reticulum; C:membrane part; C:intracellular organelle part	-	GW698102

Otpt1_0296_CZ1835_001.ab1	1	BURP domain-containing protein; unknown function	727	25	1.09E-31	62.28 %	0		-		GW697674
Otpt1_0637_CZ1855_057.ab1	1	Cytochrome b5; electron carrier	642	25	3.07E-42	73.24 %	4	C:vacuole; F:transition metal ion binding; C:endoplasmic reticulum; C:membrane part			GW697796
Otpt1_0801_CZ1855_065.ab1	1	Cytochrome b5; weakely similar to Otpt1_0637	407	25	8.91E-24	83.72 %	8	C:endoplasmic reticulum membrane; C:microsome; C:vacuole; C:cytoplasmic membrane-bounded vesicle; F:iron ion binding; F:heme binding; P:transport; C:integral to membrane			GW697852
Otpt1_0470_CZ1835_019.ab1	1	ATDAD1 defender against cell death protein; antiapoptosis	691	25	3.15E-54	93.48 %	2	C:integral to membrane; P:apoptosis			GW697729
Otpt1_0432_CZ1835_096.ab1	1	Diphenol oxidase; ; copper ion binding / oxidoreductase (IRX12/LAC4) with laccase activity; electron transport	657	25	1.26E-46	85.60 %	4	P:oxidation reduction; F:copper ion binding; F:laccase activity; P:electron transport	EC:1.10.3.2,		GW697718
Otpt1_1100_CH0101_081.ab1	1	ELF4 early flowering expressed; involved in photoperiod, red-fa-r-red and flowering time; nucleus	568	25	1.08E-52	82.48 %	0		-		GW697962
Otpt1_0514_CZ1835_064.ab1	1	ELF4-like4; early fruit	681	25	2.20E-28	68.28 %	2	F:molecular_function; P:biological_process	-		GW697747
temperate.fasta.screen.Contig36	2	PsAD1	408	3	1.63E-09	78.00 %	1	C:plastid		Otpt1_0117_HFY328_091.ab1; Otpt1_0675_CZ1855	GW697600 GW697807

										_068.ab1	
Otpt1_1371_CH0101_037.ab1	1	GPI-anchored protein	344	1	1.83E-08	79.00 %	0		-		GW698041
Otpt1_0326_CZ1835_017.ab1	1	Harpin-induced protein 1; endomembrane system	603	25	2.39E-43	59.80 %	1	P:biological_process	-		GW697682
Otpt1_1528_AM21_053.ab1	1	Hydroxyproline-rich glycoprotein; unknown function	530	13	8.20E-27	65.92 %	1	P:biological_process	-		GW698103
temperate.fasta.screen.Contig104	4	Lectin ; metal ion and glycan binding protein	1013	25	7.31E-79	69.12 %	2	F:metal ion binding; F:cation binding		Otpt1_1057_CH0101_053.ab1; Otpt1_1531_AM21_049.ab1; Otpt1_0089_HFY328_079.ab1; Otpt1_0036_HFY328_021.ab1	GW697944 GW698104 GW697592 GW697572
temperate.fasta.screen.Contig39	2	Similar to nodulin-related	730	12	3.32E-12	58.67 %	2	P:response to bacterium; P:response to salt stress	-	Otpt1_0155_HFY330_002.ab1; Otpt1_0272_HFY330_087.ab1	GW697615 GW697660
temperate.fasta.screen.Contig9	1	Metallothionein type 1 ; zinc ion binding protein	564	4	1.36E-07	79.00 %	1	F:metal ion binding		Otpt1_1232_CH0101_087.ab1	GW698018
Otpt1_1031_CH0101_037.ab1	1	Metallothionein-like protein	560	25	1.14E-19	79.08 %	1	F:metal ion binding			GW697937
temperate.fasta.screen.Contig15	1	Metallothionein-like protein type-1	509	12	6.80E-09	82.17 %	1	F:metal ion binding		Otpt1_0598_CZ1855_021.ab1	GW697782
temperate.fasta.screen.Contig10	1	Metallothionein-like protein type-3	471	25	7.55E-23	77.92 %	1	F:metal ion binding		Otpt1_0498_CZ1835_044.ab1	GW697738

temperate.fasta.screen.Contig105	4	Metallothionein-like protein type-3	446	25	6.45E-22	77.20 %	1	F:metal ion binding		Otpt1_1239_CH0101_081.ab1; Otpt1_0482_CZ1835_020.ab1; Otpt1_0644_CZ1855_051.ab1; Otpt1_0025_HFY328_010.ab1	GW698020 GW697734 GW697797 GW697567
temperate.fasta.screen.Contig68	2	Type-a response regulator; hormone mediated signal transduction and transcriptional response	728	25	5.19E-58	87.80 %	6	P:two-component signal transduction system (phosphorelay); F:DNA binding; F:two-component response regulator activity; F:protein binding; P:response to cytokinin stimulus; P:regulation of transcription, DNA-dependent	-	Otpt1_1035_CH0101_048.ab1; Otpt1_1266_CH0101_008.ab1	GW697940 GW698031
Otpt1_0097_HFY328_067.ab1	1	RS21-C6 protein; mazG nucleotide pyrophosphohydrolase domain protein	536	25	5.14E-48	84.88 %	0		-		GW697593
Otpt1_0948_CH0101_070.ab1	1	Similar to ABOUTDESOUFFLE; mitochondrial carrier protein; carnitine/acylcarnitine	526	25	8.68E-77	73.84 %	5	F:binding; C:mitochondrial inner membrane; P:transport; C:chloroplast; F:transporter activity			GW697910
temperate.fasta.screen.Contig89	3	Nicotianamine synthase	813	25	1.28E-91	73.20 %	3	F:transferase activity; F:nicotianamine synthase activity; P:nicotianamine biosynthetic process	-	Otpt1_0292_CZ1835_007.ab1; Otpt1_0481_CZ1835_022.ab1; Otpt1_0429_CZ1835_085.ab1	GW697672 GW697733 GW697716
Otpt1_0814_CZ1855_068.ab1	1	Pollen-specific protein	636	25	6.22E-64	69.96 %	0		-		GW697855

Otpt1_0546_CZ1835_091.ab1	1	Protein kinase c inhibitor; zinc binding protein	626	25	2.11E-53	86.72 %	2	F:zinc ion binding; F:catalytic activity			GW697766
Otpt1_0073_HFY328_057.ab1	1	SHOOT1 protein; chloroplast thylakoid membrane, protein binding	617	25	1.17E-64	76.12 %	2	F:protein binding; C:plastid			GW697585
Otpt1_0476_CZ1835_028.ab1	1	Short-chain dehydrogenase reductase family protein	628	25	1.92E-70	77.20 %	2	F:oxidoreductase activity; P:defense response			GW697732
temperate.fasta.screen.Contig82	3	Specific tissue protein 2; 6 repeats of 84 bp	819	9	2.24E-20	45.11 %	0		-	Otpt1_1184_CH0101_049.ab1; Otpt1_0796_CZ1855_075.ab1; Otpt1_0398_CZ1835_077.ab1	GW698005 GW697850 GW697706
temperate.fasta.screen.Contig103	4	Specific tissue protein 2; 3 repeats of 78 bp	737	16	4.12E-38	60.31 %	0		-	Otpt1_0945_CH0101_074.ab1; Otpt1_1464_AM21_016.ab1; Otpt1_0265_HFY330_070.ab1; Otpt1_0420_CZ1835_068.ab1	GW697909 GW698076 GW697654 GW697711
temperate.fasta.screen.Contig35	2	Specific tissue protein 2; 3 imperfect repeats of 78 bp	721	19	3.33E-25	56.84 %	0		-	Otpt01_0008_CZ1321_058.ab1; Otpt1_1454_CH0101_082.ab1	GW698114 GW698069
temperate.fasta.screen.Contig5	1	Specific tissue protein 2	664	9	1.87E-13	45.78 %	0		-	Otpt1_0770_CZ1855_034.ab1	GW697841
Otpt1_0899_CH0101_038.ab1	1	Hypothetical protein; endomembrane system	567	10	5.96E-27	75.30 %	1	P:biological_process	-		GW697889

Otpt1_0664_CZ1855_071.ab1	1	Hypothetical protein; SWIB complex BAF60b domain-containing protein	689	3	7.51E-08	62.00 %	0		-		GW697802
Otpt1_1195_CH0101_056.ab1	1	Hypothetical protein; Endomembrane system	770	11	8.96E-35	73.64 %	1	P:biological_process			GW698008
Otpt1_0256_HFY330_080.ab1	1	Hypothetical protein	636	25	2.88E-69	77.60 %	2	F:molecular_function; P:biological_process	-		GW697651
Otpt1_0698_CZ1855_086.ab1	1	Hypothetical protein; chloroplast; unknown function;	600	7	2.02E-10	72.43 %	1	P:biological_process	-		GW697818
temperate.fasta.screen.Contig51	2	Hypothetical protein; unknown function	1082	25	3.56E-143	73.24 %	2	C:plastid; C:membrane		Otpt1_1106_CH0101_092.ab1; Otpt1_1093_CH0101_093.ab1	GW697964 GW697958
Otpt1_0858_CH0101_006.ab1	1	Hypothetical protein; unknown funcion	445	7	1.35E-11	79.71 %	2	C:plastid; C:membrane			GW697871
Otpt1_0270_HFY330_091.ab1	1	Hypothetical protein; unknown function	685	20	8.18E-23	60.10 %	1	C:plastid	-		GW697658
*Otpt1_0918_CH0101_064.ab1	1	Hypothetical protein; chloroplast thylakoid membrane; drought stress response	679	20	7.49E-29	82.00 %	1	C:plastid	-		GW697896
Otpt1_0445_CZ1835_013.ab1	1	Hypothetical protein; unknown function	732	25	1.20E-50	72.32 %	1	P:biological_process	-		GW697721

Otpt1_0236_HFY330_056.ab1	1	Hypothetical protein; unknown function	672	25	2.05E-55	70.72 %	3	C:plastid; F:molecular_function; P:biological_process	-		GW697644
Otpt1_1484_AM21_030.ab1	1	Hypothetical protein; unknown function	531	1	3.66E-06	62.00 %	0		-		GW698086
temperate.fasta.screen.Contig7	1	Hypothetical protein	624	0			0		-	Otpt1_1064_CH0101_060.ab1	GW697946
temperate.fasta.screen.Contig72	2	Hypothetical protein	624	0			0		-	Otpt1_0674_CZ1855_070.ab1; Otpt1_0761_CZ1855_048.ab1	GW697806 GW697839
Otpt1_0274_HFY330_083.ab1	1	Hypothetical protein	645	1	9.44E-06	60.00 %	0		-		GW697662
Otpt1_0704_CZ1855_013.ab1	1	Hypothetical protein	450	12	1.17E-07	86.50 %	0		-		GW697822
Otpt1_1467_AM21_010.ab1	1	Hypothetical protein	725	25	8.22E-48	75.72 %	1	C:cytoplasmic membrane-bounded vesicle	-		GW698079
temperate.fasta.screen.Contig78	2	Hypothetical protein	398	0			0		-	Otpt1_1156_CH0101_035.ab1; Otpt1_1242_CH0101_094.ab1	GW697994 GW698021
Otpt1_1255_CH0101_007.ab1	1	Hypothetical protein , similar serine acetyltransferase and to Salmonella genome	475	25	1.08E-37	74.72 %	0		-		GW698028
temperate.fasta.screen.Contig112	6	Hypothetical protein 71 Cicer	578	1	2.38E-18	76.00 %	0		-	Otpt1_0195_HFY330_046.ab1; Otpt1_0336_CZ1835	GW697634 GW697687 GW697610

										_020.ab1; Otpt1_0140_HFY330_005.ab1; Otpt1_0503_CZ1835_034.ab1; Otpt1_0174_HFY330_026.ab1; Otpt1_1121_CH0101_016.ab1	GW697742 GW697625 GW697974
temperate.fasta.screen.Contig113	6	Hypothetical protein 71 Cicer	555	1	4.72E-18	76.00 %	0		-	Otpt1_0936_CH0101_071.ab1; Otpt1_1491_AM21_020.ab1; Otpt1_0680_CZ1855_089.ab1; Otpt1_0854_CH0101_012.ab1; Otpt1_1173_CH0101_063.ab1; Otpt1_0501_CZ1835_038.ab1	GW697903 GW698090 GW697809 GW697869 GW698001 GW697740
Otpt1_1445_CH0101_094.ab1	1	Cyclin-like F-box (weak similarity)	723	0			0		-		GW698064
Otpt01_0004_CZ1321_057.ab1	1	F-box family protein	246	7	1.04E-16	66.00 %	0		-		GW698113
temperate.fasta.screen.Contig75	2	F-box family protein	781	4	6.62E-41	56.00 %	2	F:molecular_function; P:biological_process	-	Otpt1_0702_CZ1855_082.ab1; Otpt1_0703_CZ1855_015.ab1	GW697820 GW697821
Otpt1_0795_CZ1855_077.ab1	1	hit soybean and medicago genomes	754	1			0		-		GW697849
Otpt1_0993_CH0101_004.ab1	1	hit soybean and medicago genomes	641	0			0		-		GW697922

temperate.fasta.screen.Contig11	1	hit soybean and populus genomes	689	0	0	-	Otpt1_0931_CH0101_079.ab1	GW697901
temperate.fasta.screen.Contig14	1	hit soybean and populus genomes	668	0	0	-	Otpt1_1448_CH0101_090.ab1	GW698066
Otpt1_1152_CH0101_043.ab1	1	hit soybean and populus genomes	666	0	0	-		GW697991
Otpt1_1251_CH0101_015.ab1	1	hit soybean and populus genomes	585	0	0	-		GW698026
temperate.fasta.screen.Contig100	4	hit to soybean genome	544	0	0	-	Otpt1_1096_CH0101_087.ab1; Otpt1_1101_CH0101_096.ab1; Otpt1_0207_HFY330_036.ab1; Otpt1_0142_HFY330_001.ab1	GW697959 GW697963 GW697636 GW697611
temperate.fasta.screen.Contig74	2	hit to soybean genome	637	0	0	-	Otpt1_0237_HFY330_054.ab1; Otpt1_0038_HFY328_017.ab1	GW697645 GW697574
temperate.fasta.screen.Contig44	2	hit to soybean genome	473	0	0	-	Otpt1_1499_AM21_041.ab1; Otpt1_0538_CZ1835_072.ab1	GW698092 GW697760
temperate.fasta.screen.Contig73	2	hit to soybean genome	421	0	0	-	Otpt1_0444_CZ1835_015.ab1; Otpt1_1118_CH0101_005.ab1	GW697720 GW697971
Otpt1_1540_AM21_052.ab1	1	hit to soybean genome	691	0	0	-		GW698108

Otpt1_1024_CH0101_047.ab1	1	hit to soybean genome	619	0	0	-		GW697932
Otpt1_0362_CZ1835_063.ab1	1	hit to soybean genome	605	0	0	-		GW697694
Otpt1_0977_CH0101_013.ab1	1	hit to soybean genome	595	0	0	-		GW697916
Otpt1_0168_HFY330_017.ab1	1	Mitochondrial protein	229	0	0	-		GW697623
		Unknown						
temperate.fasta.screen.Contig53	2	NA	701	0	0	-	Otpt1_1451_CH0101_088.ab1; Otpt1_0567_CZ1855_011.ab1	GW698067 GW697774
Otpt1_1451_CH0101_088.ab1	1	NA	701	0	0	-		
Otpt1_0567_CZ1855_011.ab1	1	NA	533	0	0	-		
temperate.fasta.screen.Contig66	2	NA	754	0	0	-	Otpt1_0422_CZ1835_095.ab1; Otpt1_0360_CZ1835_036.ab1	GW697712 GW697692
temperate.fasta.screen.Contig28	1	NA	655	0	0	-	Otpt1_0158_HFY330_029.ab1	GW697617
temperate.fasta.screen.Contig19	1	NA	635	0	0	-	Otpt1_0466_CZ1835_027.ab1	GW697726

temperate.fasta.screen.Contig58	2	NA	570	0	0	-	Otpt1_0739_CZ1855_032.ab1; Otpt1_0067_HFY328_038.ab1	GW697833 GW697581
temperate.fasta.screen.Contig69	2	NA	379	0	0	-	Otpt1_1120_CH0101_001.ab1; Otpt1_1163_CH0101_044.ab1	GW697973 GW697997
Otpt1_0352_CZ1835_046.ab1	1	NA	706	0	0	-		GW697690
Otpt1_0611_CZ1855_047.ab1	1	NA	673	0	0	-		GW697789
Otpt1_0250_HFY330_067.ab1	1	NA	589	0	0	-		GW697650
Otpt1_0028_HFY328_004.ab1	1	NA	548	0	0	-		GW697568
Otpt1_1521_AM21_063.ab1	1	NA	469	0	0	-		GW698100
Otpt1_0782_CZ1855_051.ab1	1	NA	440	0	0	-		GW697845
Otpt1_0839_CZ1855_082.ab1	1	NA	422	0	0	-		GW697863
Otpt1_0302_CZ1835_014.ab1	1	NA	414	0	0	-		GW697675
Otpt1_0886_CH0101_045.ab1	1	NA	413	0	0	-		GW697881

Otpt1_1243_CH0101_092.ab1	1	NA	410	0	0	-	GW698022
Otpt01_0003_CZ1321_042.ab1	1	NA	399	0	0	-	GW698112
Otpt1_1060_CH0101_064.ab1	1	NA	370	0	0	-	GW697945
Otpt1_0898_CH0101_040.ab1	1	NA	327	0	0	-	GW697888
Otpt1_0892_CH0101_035.ab1	1	NA	299	0	0	-	GW697883
Otpt1_0461_CZ1835_006.ab1	1	NA	275	0	0	-	GW697723
Otpt1_0286_HFY330_084.ab1	1	NA	273	0	0	-	GW697668
Otpt1_0062_HFY328_046.ab1	1	NA	234	0	0	-	GW697578
Otpt1_0173_HFY330_028.ab1	1	NA	201	0	0	-	GW697624
Otpt1_0012_HFY328_013.ab1	1	NA	193	0	0	-	
Otpt1_0187_HFY330_039.ab1	1	NA	171	0	0	-	

A vertical line before the first column indicates that adjacent contigs have similar sequence.

* An asterisk indicates there is a similar contig in the temperate *Oxytropis* enriched library, and that the gene is a potential false positive.

8.5 Appendix 5: Supplementary Tables S3.5 to S3.14 reporting sequence divergence values between unique genes of gene families from *Oxytropis* transcriptomes.

Supplementary Table S3.5 Estimates of divergence between sequences of unique genes from the arctic-enriched and the temperate enriched annotated as pathogenesis related protein; class 10.
The number of base differences per site between sequences are shown. This pairwise distance analysis involved 12 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 818 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Kumar et al. 2008).

arctic.fasta.screen.Contig2											
arctic.fasta.screen.Contig36	**0.0846**										
arctic.fasta.screen.Contig9	0.1248	0.1481									
arctic.fasta.screen.Contig13	0.1156	0.1109	0.0741								
arctic.fasta.screen.Contig34	0.117	0.1048	0.115	0.0863							
arctic.fasta.screen.Contig39	0.0876	0.1071	0.096	0.0747	0.0639						
arctic.fasta.screen.Contig14	0.1002	0.0888	0.116	0.0536	0.0748	0.0644					
arctic.fasta.screen.Contig43	0.0893	**0.0962**	0.1145	0.0521	**0.0497**	0.0405	0.0342				
arctic.fasta.screen.Contig18	0.1139	0.1263	0.0755	0.0928	0.0991	0.0895	0.074	0.106			
arctic.fasta.screen.Contig61	0.0962	0.0994	0.1151	**0.0493**	0.0533	0.0459	**0.0324**	0.0115	0.1095		
temperate.fasta.screen.Contig54	0.0929	0.1104	0.0942	0.0782	0.0524	**0.0323**[a]	0.0615	0.0363	0.1023	0.0389	
Otpt1_0104_HFY328_072.ab1	0.2785	0.2864	**0.2601**	0.2729	0.2676	0.2679	0.2656	0.2837	0.2636	0.2905	0.2805

Supplementary Table S3.6 Estimates of divergence between sequences of unique genes from the arctic-enriched and the temperate enriched annotated as defensin protein; gamma-thionin/defensin, PDF1.
The number of base differences per site between sequences are shown. This pairwise distance analysis involved 9 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 992 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Kumar et al. 2008).

arctic.fasta.screen.Contig15								
arctic.fasta.screen.Contig56	**0.0417**							
arctic.fasta.screen.Contig38	0.1118	**0.1016**						
Oapa1_1097_CZ1898_015.ab1	0.1102	0.1164	**0.0463**					
arctic.fasta.screen.Contig50	0.4729	0.4379	0.1224	0.0782				
arctic.fasta.screen.Contig51	0.456	0.4244	0.1148	**0.0769**	**0.0291**			
arctic.fasta.screen.Contig26_	0.4615	0.4294	0.0944	0.1102	**0.055**	0.0558		
arctic.fasta.screen.Contig41	0.4733	0.4724	0.0922	0.1162	0.0562	0.0594	**0.0249**	
arctic.fasta.screen.Contig24	0.4492	0.4492	0.1466	0.1294	0.0564	0.0571	**0.0415**	0.0499

Supplementary Table S3.7 Estimates of divergence between sequences of unique genes from the arctic-enriched and the temperate enriched annotated as cold-regulated or drought induced protein (cold dehydrin).
The number of base differences per site between sequences are shown. This pairwise distance analysis involved 14 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1342 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Kumar et al. 2008).

arctic.fasta.screen.Contig12													
arctic.fasta.screen.rcContig21	0.1534												
arctic.fasta.screen.Contig47	0.1597	0.0604											
arctic.fasta.screen.Contig25	0.1691	0.0757	0.0959										
arctic.fasta.screen.Contig23	0.2164	0.1134	0.0952	**0.0564**									
arctic.fasta.screen.Contig7	0.1413	0.0495	**0.042**	0.0888	0.0924								
arctic.fasta.screen.Contig27	0.2584	0.076	0.102	0.1061	0.1522	0.0938							
Oapa1_0752_CZ1898_047.ab1	0.2584	0.0812	0.0969	0.1114	0.1504	0.0922	**0**						
arctic.fasta.screen.Contig44	0.1448	0.0351	0.0455	0.08	0.1067	0.0367	0.0748	0.0775					
arctic.fasta.screen.Contig59	**0.13**	0.0348	0.0728	0.0594	0.1121	**0.0363**	0.0814	0.0845	0.0327				
Oapa1_0429_CZ1835_063.ab1	0.1642	**0.029**	0.0564	0.0625	0.1206	0.0385	0.0453	0.049	0.0293	0.0067			
temperate.fasta.screen.Contig111	0.4277	0.3878	0.4093	0.3932	0.4134	0.3881	0.4201	0.4316	0.383	0.378	0.3801		
temperate.fasta.screen.Contig22	0.4205	0.375	0.4047	0.3948	0.4141	0.389	0.4264	0.4415	0.3904	**0.3771**	0.3824	**0.0431**	
Otpt1_0137_HFY330_009.ab1	0.562	0.4783	0.5625	0.5581	0.5512	0.5476	0.6346	0.6346	0.5738	0.5606	0.5534	**0.3443**	0.3504

Supplementary Table S3.8 Estimates of divergence between sequences of unique genes from the arctic-enriched and the temperate enriched annotated as chlorophyll a b-binding proteins; light-harvesting protein of photosystem I and II.
The number of base differences per site between sequences are shown. This pairwise distance analysis involved 14 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1155 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Kumar et al. 2008).

Otpt1_0074_HFY328_055.ab1											
temperate.fasta.screen.Contig45	0.5662										
temperate.fasta.screen.Contig110	0.584	0.6161									
temperate.fasta.screen.Contig99	0.5517	0.5779	0.6114								
temperate.fasta.screen.Contig96	0.5714	0.5644	0.5867	0.5495							
temperate.fasta.screen.Contig92	0.5657	0.5947	0.5923	0.566	0.5134						
temperate.fasta.screen.Contig21	0.5408	0.5406	0.5983	0.5464	0.4338	0.2256					
temperate.fasta.screen.Contig119	0.573	0.5843	0.591	0.56	0.4805	0.2045	**0.02**				
Otpt1_1153_CH0101_041.ab1	0.5249	0.5201	0.5204	0.5268	0.4111	0.3422	0.3371	0.3329			
temperate.fasta.screen.Contig98	0.5699	0.5734	0.5673	0.5573	0.476	0.3593	0.2926	0.3397	0.3075		
Otpt1_1115_CH0101_011.ab1	0.5744	0.5471	0.5718	0.5288	0.4751	0.3561	0.251	0.3285	0.2617	**0.0119**	
Otpt1_0522_CZ1835_052.ab1	0.5697	0.5582	0.5836	0.5786	0.4491	0.372	0.288	0.3129	0.3177	**0.0232**	0.0176

Supplementary Table S3.9 Estimates of divergence between sequences of unique genes from the arctic-enriched and the temperate enriched annotated as non-specific lipid-transfer protein (LTP).
The number of base differences per site between sequences are shown. This pairwise distance analysis involved 5 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 774 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Kumar et al. 2008).

temperate.fasta.screen.Contig18		
temperate.fasta.screen.Contig2	0.0603	
temperate.fasta.screen.Contig106	**0.0601**	0.0634

temperate.fasta.screen.Contig102	0.1718	**0.0522**	0.1293		
temperate.fasta.screen.Contig31	**0.101**	0.1107	0.1312	0.1677	
Oapa1_1518_AM23_079.ab1_LTP	0.4589	0.4309	0.4882	0.4818	**0.3992**

Supplementary Table S3.10 Estimates of divergence between sequences of unique genes from the arctic-enriched and the temperate enriched annotated as ripening related protein (RRP); bet v 1; major latex-like protein.
The number of base differences per site between sequences are shown. This pairwise distance analysis involved 7 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 842 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Kumar et al. 2008).

temperate.fasta.screen.Contig48						
temperate.fasta.screen.Contig32	0.2116					
temperate.fasta.screen.Contig81	**0.2045**	0.0923				
temperate.fasta.screen.Contig101	0.1902	**0.0917**	0.0581			
temperate.fasta.screen.Contig8	0.1944	0.0977	**0.0475**	0.0514		
temperate.fasta.screen.Contig17	0.1922	0.128	0.0741	0.0973	**0.045**	
temperate.fasta.screen.Contig93	0.2036	0.1241	0.096	0.109	0.0477	**0.0215**

Supplementary Table S3.11 Estimates of divergence between sequences of unique genes from the arctic-enriched and the temperate enriched annotated as aluminum induced; embryonic abundant protein precursor-like protein (ADR6).
The number of base differences per site between sequences are shown. This pairwise distance analysis involved 3 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 846 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Kumar et al. 2008).

temperate.fasta.screen.Contig52		
temperate.fasta.screen.Contig91	**0.0322**	
temperate.fasta.screen.Contig95	0.0888	0.0823

Supplementary Table S3.12 Estimates of divergence between sequences of unique genes from the arctic-enriched and the temperate enriched annotated as specific tissue protein 2 (STP).
The number of base differences per site between sequences are shown. This pairwise distance analysis involved 3 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1035 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Kumar et al. 2008).

temperate.fasta.screen.Contig103			
temperate.fasta.screen.Contig35	**0.3029**		
temperate.fasta.screen.Contig5	0.358	0.1727	
temperate.fasta.screen.Contig82	0.3561	**0.1609**	**0.041**

Supplementary Table S3.13 Estimates of divergence between sequences of unique genes from the arctic-enriched and the temperate enriched annotated as vegetative storage protein; acid phosphatase (VSP).
The number of base differences per site between sequences are shown. This pairwise distance analysis involved 3 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 739 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Kumar et al. 2008).

Otpt1_0031_HFY328_029.ab1		
Otpt1_1461_AM21_005.ab1	**0.2406**	
temperate.fasta.screen.Contig80	0.5007	0.4856

Supplementary Table S3.14 Estimates of divergence between sequences of unique genes from the arctic-enriched and the temperate enriched annotated as Metallothionein-like protein type-1.

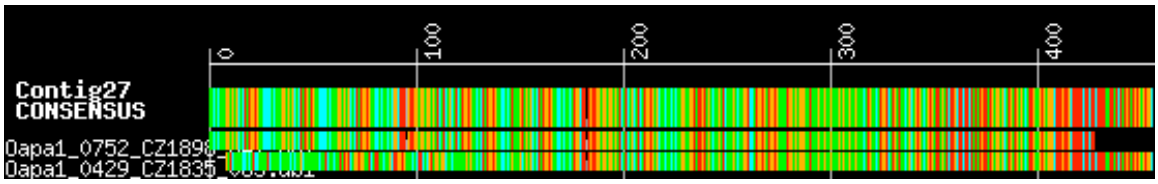
The number of base differences per site between sequences are shown. This pairwise distance analysis involved 3 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 615 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Kumar et al. 2008).

Otpt1_1031_CH0101_037.ab1		
temperate.fasta.screen.Contig15	0.3707	
temperate.fasta.screen.Contig9	**0.3772**	**0.0258**

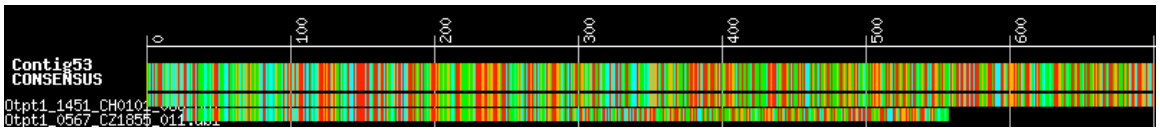
Supplementary Table S3.15 Estimates of divergence between sequences amplified fragments from genomic DNA with primers used for real-time RT-PCR for two genes of the pathogenesis-related class 10 (PR-10) genes.
The number of base differences per site between sequences are shown. This pairwise distance analysis involved 9 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 674 positions in the final dataset, including a 134 region of intron sequence. Evolutionary analyses were conducted in MEGA5 (Kumar et al. 2008).

OaPR10_61_65e_HM107135								
OmPR10_61_91b_HM107139	0.1072							
OsPR10_61_44g_HM107137	0.1111	0.0172						
OsPR10_61_88f_HM107138	0.11	0.0153	0					
OcjPR10_61_71a_HM107136	0.1342	0.1142	0.1196	0.1102				
OcjPR10_13.36_71a_HM107142	0.1324	0.0633	0.0576	0.0481	0.153			
OmPR10_13.36_91b_HM107143	0.0789	0.1098	0.1139	0.1037	0.1059	0.0497		
OaPR10_13.36_46_HM107140	0.0866	0.1164	0.1243	0.1181	0.1286	0.0524	0.0303	
OsPR10_13.36_44g_HM107141	0.0887	0.1426	0.1495	0.1418	0.1524	0.0748	0.0231	0.0329

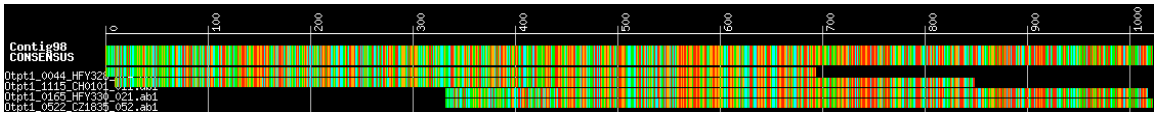
a)



b



temperate.contig98.png



8.6 Appendix 6: Supplementary Figure S3.1 Images for contigs involving suspected misassembled sequence, for sequence assembly of arctic-enriched and temperate enriched plantlets *Oxytropis* cDNA libraries.

a) contig27 from arctic-enriched library, including Oapa1_0752_CZ1898_047.ab1 and Oapa1_0429_CZ1835_063.ab1 ESTs; b) contig53 from the temperate-enriched library, including Otpt1_1451_CH0101_088 and Otpt1_0567_CZ1855_011 ESTs; c) contig98 from temperate-enriched library, including Otpt1_1115_CH0101_011.ab1 and Otpt1_0522_CZ1835_052.ab1 ESTs.

8.7 Appendix 7: Supplementary Table S4.1 *Oxytropis* seed sources used for genomic DNA extraction and PCR amplification.

Oxytropis species	Ploidy level	Locality	Plantlet No	Collector
From the Arctic regions				
O. arctobia Bunge	2n = 2x = 16 (Elven 2007)	Nanisivik, Baffin Island (Nunavut, Canada, 73.03° N 84.55° W)	46, 46e, 65b, 65c, 65e	Dr. Susan Aiken, Canadian Museum of Nature, Ottawa
O. arctobia		Sarcpa Lake, Melville Peninsula (Nunavut, Canada, 68.52° N 83.27° W)	43b, 57a	Dr. Danielle Prévost, Agriculture and Agri-Food Canada, Ste-Foy (Prevost et al. 1987)
O. maydelliana Trautv.	(2n = 12X = 96, (Holmen 1962)	Apex, Baffin Island (Nunavut, Canada, 63.72° N 68.48° W)	20g, 40e, 47b	Dr. Susan Aiken, Canadian Museum of Nature, Ottawa
O. maydelliana		Sarcpa Lake, Melville Peninsula (Nunavut, Canada, 68.52° N 83.27° W)	91b. 101c, 101b	Dr. Danielle Prévost, Agriculture and Agri-Food Canada, Ste-Foy (Prevost et al. 1987)
From the temperate regions				
O. campestris (L.) DC subsp. *johannensis* (Fernald) Blondeau & Gervais	2n = 6X = 48, (Ledingham 1960)	Ile d'Orléans, Quebec City (Québec, Canada, 45.85° N 71.05° W)	12, 17, 71a, 87a, 98a	Annie Archambault
O. splendens Douglas	2n = 2X = 16 (Ledingham 1957)	Southern Alberta, Canada	CN105143	Plant Gene Resources of Canada (Ottawa) (CN105143)
O. splendens		Longview, (Southern Alberta, Canada, 50.53° N 114.23° W)	23d, 88f, 99a	Wild About Flowers, native plant nursery
O. splendens		Cochin, (Southern, Saskatchewan, Canada, 53.08° N 108.34° W)	44, 44h, 44g, 54a	Prairie Garden Seeds, native plant nursery
From a nearly arid subtropical region				
O. lambertii Pursh	2n = 6X = 48 (Ledingham 1957)	Ocate, New Mexico (USA, 36.21° N 105.0° W)	18, 18f, 41	Dr. Tracy Sterling New Mexico State University, Las Cruces (Kulshreshtha et al. 2004)

8.8 Appendix 8: Supplementary Table S4.2 Sequence, location and direction for primers used in PCR amplification of selected genes from four *Oxytropis* species gDNA. Primers were named according to the cDNA library of origin, target contig and amplification direction.

Primer name	Direction	Primer sequence (5' to 3')	Primer location [a]	Annealing temperature in °C for Amplitaq Gold[b]	Annealing temperature in °C for iProof [b]
KS-dehydrin					
arct_cold_all_1R	Reverse	AAGCAACAAAGCCCTCACTTC	506-526	62 to 66	
arct_cold47_1F	Forward	GCAAACCACATCCAAAACCAAAA	15-37	65 to 67	71
arct_cold47_2R	Reverse	GGTCTCATCATGCTCCTGCAACT	556-578	65 to 67	71
arct_cold47_6R	Reverse	AACAAGAAACACGCATTTTCTC	663-686	60 to 61	
arct_cold59_1F	Forward	GGATCAAAACCACATCCAAAATG	06-28	62	
arct_cold59_3F	Forward	ACACTTCACGTAGGAGGCAACAA	96-118	62 to 65	69
arct_cold59_4R	Reverse	CACAAACCAATCTATCCGCAAAT	631-653	59 to 62	69
arct_cold59_5F	Forward	TGAACCACAGAAAGCAGAGCA	246-266	62 to 63	
arct_cold23_1F	Forward	CAAACCACATCCCAAACCAATC	06-27	59 to 65	
arct_cold27_1F	Forward	ATGAACCCCACAAAGGAGAGCAC	49-71	61 to 65	71
arct_cold27_3F	Forward	CATCATGGTGAACACTATGGTG	118-139	56 to 61	71
arct_cold44_1F	Forward	GTCAGGAATCATTAACGAGA	52-71	56 to 61	
arct_cold25_3F	Forward	GGAGAGCACTATGGTAACCAC	17-37	59 to 66	
arct_cold25_1F	Forward	AACCCTTCACGTAGGAGRCCACA	90-112	60 to 66	68
arct_cold25_2R	Reverse	GGTGGGGTTTCACAAWCCAATCT	613-635	60 to 66	71
Pathogenesis related class 10					
PR10_18_3F	Forward	TGGCCCCGGAACCACCAAGAAAC	180-202	60 to 65	
PR10_18_2R	Reverse	AGTCCACACAACATGCATGGAAC A	552-575	60 to 65	
PR10_34_1F	Forward	CCGTTGAAGGAAATGGTGGTCCC	197-219	57 to 59	
arct_PR10_34_4R	Reverse	AAACGTAACCCTCGATGGCCT	503-523	59 to 63	68
arct_PR10_36_1F	Forward	CCTACCAACCATGGGAATCTYCA	47-69	61; 71	
arct_PR10_36_3F	Forward	CACATTTGAGCAAGAAACCACCTC	68-91	59 to 63	71
arct_PR10_36_2R	Reverse	ACATAACCCTCGATGGCCTTGAA	486-508	61	71
arct_PR10_36_4R	Reverse	TTAGGATTGGCCAAAACATAACC	501-523	61 to 64	71
arct_PR10_61_1F	Forward	CAACACAACACAACWGCCAACAA	18-40	63 to 66	
arct_PR10_61_3F	Forward	CATGGGAATCTTCACATTTGAGC	87-109	NA	71
arct_PR10_61_2R	Reverse	CCACACAACAAGCATGCAACAAG	603-625	62 to 65	
arct_PR10_61_4R	Reverse	TTGATGAAGCGTGATGTAAAAAC A	699-722	63 to 65	71
arct_PR10_61_5F	Forward	TCGTTGAAGGAAATGGTGGCCCT	136-158	65	
arct_PR10_61_11F_only	Forward	CCTCTACTGTTGCTCCTGCCAA	40-61	63 to 64	
arct_PR10_61_16R_only	Reverse	AAACGTAACCCTCGATAGCCT	521-541	63 to 66	

arct_PR10_61_13F_only	Forward	GTCATCAAGAGTGTTGAAATC	196-216	63	
arct_PR10_61_12R_only	Reverse	CACATTGAGCTTTCCAATAGAGC	292-314	64 to 65	
arct_PR10_13_1F	Forward	CTCCAAGTCATCAATTCATCACCAC	50-74	63 to 64	71
arct_PR10_13_2R	Reverse	TTGGCCAAAACGTAACCCTCGAT	522-544	60 to 64	71
arct_PR10_13_5F	Forward	AAGTAACTATGGGTGTTTTCAC	76-97	64	
arct_PR10_13_4R	Reverse	TTGGTATTTGACAGTGAGCTTCC	424-446	64	
PR10_Fgroup_3R	Reverse	CCTTCCTTAACTTCSTTTTCAA	-	59 to 61	
Ripening-related proteins					
temp_ripen_101_1F	Forward	CWGAACTTGGCATCAAATCACCA	85-107	NA	68
temp_ripen_101_2R	Reverse	CAAGCATGATTYTATACATGCCAAA	614-638	NA	68
temp_ripen_101_3F	Forward	GCTGCCAAATTCTTCAACCTCTT	108-130	60 to 66	68
temp_ripen_101_5R	Reverse	TTACAAGAACWTCACCAAATCCA	566-588	60 to 66	68
temp_ripen32_1F	Forward	CATTAGTCATGGTTCTCGCTGG	61-82	60 to 65	
temp_ripen32_2R	Reverse	TGAAGAAGAACCCACAGATGCA	653-674	60 to 65	
temp_ripen_93_1F	Forward	CATGGTACTCGCAGGGAAACTCA	68-90	NA	71
temp_ripen_93_2R	Reverse	TTGCACACATGACGGAGAAGGAT	664-686	62 to 64	71
temp_ripen_93_3F	Forward	GCATCAAATCACCAGCTGCAAAG	103-125	62 to 64	
temp_ripen_93_5F	Forward	CCTCTTTGCAAAAGAACTTCAC	134-155	62 to 64	
temp_ripen_93_4R	Reverse	TCTCCAGACATTTGTGGACCTTGC	259-282	62 to 64	
lhcbI					
temp_lhcbI_119_1F	Forward	TGGGAGCTTCCAAGTTCACAATG	152-174	64 to 67	70
temp_lhcbI_119_2R	Reverse	CTACTCTCACTTTCCGGGGACGA	857-859	64 to 67	70
lhcaIII					
temp_lhcaIII_110_3F	Forward	AAATTCAGGAGGGAGGAAAAAGC	57-79	NA	68
temp_lhcaIII_110_1F	Forward	CATGGCTGCACAAGCTCTGGTAT	79-101	60 to 68	71
temp_lhcaIII_110_2R	Reverse	GTGCTCAGCAAATCCCATGAGTG	576-598	60 to 68	71
temp_lhcaIII_110_4R	Reverse	GCATCCTTACCAAAACCAAGAGG	711-738	NA	68

[a] Numbering is based on the position of the primer 3' nucleotide relative to the target contig sequence (Archambault and Strömvik 2011).

[b] Compatible annealing temperatures are indicated for amplification reactions using Amplitaq Gold (Applied Biosystems), or using iProof Polymerase (BioRad). See method section for more details. Primers lhcaIII_110_1F with lhcaIII_110_2R; lhcbI_119_1F with lhcbI_119_2R; arct_cold27_1F with arct_cold25_2R; arct_cold59_3F with arct_cold59_4R and PR10_13_1F with PR10_13_2R required iProof polymerase for a strong amplification.

8.9 Appendix 9: Supplementary Table S4.3 Size of introns for the orthologs and paralogs for PR-10, ripening-related proteins, KS-dehydrins, and *lhcaIII* isolated from *Oxytropis arctobia* (Oa), *O. maydelliana* (Om), *O. campestris johannensis* (Ocj) *O. splendens* (Os), and *O. lambertii* (Ol).

Sequence name	Accession number	Oxytropis species	Intron length (bp)
PR-10			
PR-10_f_om_Prevost_sn_101c	HQ731827	Om	154
PR-10_f_om_Aiken_sn_40e	HQ731830	Om	154
PR-10_f_om_Aiken_sn_20g	HQ731828	Om	154
PR-10_f_os_PGS_44	HQ731826	Os	100
PR-10_13_ol_Sterling_s41	HQ731809	Ol	94
PR-10_13_oa_Aiken_sn_46_76bp	HQ731802	Oa	76
PR-10_13_oa_Aiken_sn_46_99bp	HQ731808	Oa	99
PR-10_13_os_PGS_44	HQ731803	Os	108
PR-10_13_os_PGS_44g	HQ731804	Os	108
PR-10_13_ocj_AA300_71a	HQ731810	Ocj	91
PR-10_13_ocj_AA300_12	HQ731811	Ocj	91
PR-10_13_os_WAF_88f	HQ731812	Os	95
PR-10_18_os_PGS_44g	HQ731834	Os	134
PR-10_18_os_WAF_88f	HQ731833	Os	134
PR-10_18_oa_Aiken_sn_46	HQ731832	Oa	138
PR-10_18_oa_Prevost_sn_43b	HQ731831	Oa	138
PR-10_18_ocj_AA300_71a	HQ731835	Ocj	148
PR-10_18_oa_Aiken_sn_65e	HQ731836	Oa	117
PR-10_18_oa_Aiken_sn_46e	HQ731829	Oa	138
PR-10_61_oa_Aiken_sn_46e	HQ731823	Oa	117
PR-10_61_oa_Aiken_sn_65e	HQ731822	Oa	117
PR-10_61_oa_Aiken_sn_65c	HQ731821	Oa	117
PR-10_61_om_Prevost_sn_91b	HQ731813	Om	81
PR-10_61_om_Aiken_sn_20g	HQ731824	Om	115
PR-10_61_om_Aiken_sn_40e	HQ731825	Om	97
PR-10_61_os_CN_105143_sn	HQ731815	Os	91
PR-10_61_ocj_AA300_71a_82bp	HQ731820	Ocj	82
PR-10_61_ocj_AA300_71a_86bp	HQ731819	Ocj	86
PR-10_61_os_PGS_44g	HQ731818	Os	91
PR-10_61_os_PGS_44h	HQ731817	Os	91
PR-10_61_os_PGS_44	HQ731816	Os	91
Ripening-related proteins			
ripening_p_ocj_AA300_71a	HQ731891	Ocj	145
ripening_p_oa_Aiken_sn_65c	HQ731892	Oa	145
ripening_p_os_CN_105143_sn	HQ731894	Os	145

ripening_p_om_Aiken_sn_20g	HQ731893	Om	145
ripening_32_om_Aiken_sn_20g	HQ731895	Om	350
ripening_32_os_PGS_44g	HQ731896	Os	352
ripening_32_om_Aiken_sn_94b	HQ731898	Om	352
ripening_32_om_Prevost_sn_101c	HQ731897	Om	350
ripening_101_ocj_Prevost_sn_101c	HQ731899	Ocj	798
ripening_101_oa_Aiken_sn_46e	HQ731901	Oa	830
ripening_101_om_Aiken_sn_40e	HQ731900	Om	unknown
ripening_101_os_PGS_44g	HQ731902	Os	798
ripening_93_oa_Aiken_sn_65e	HQ731904	Oa	715
ripening_93_oa_Aiken_sn_46e	HQ731903	Oa	715
ripening_93_os_WAF_23d	HQ731905	Os	721
ripening_93_om_Aiken_sn_40e	HQ731906	Om	715
ripening_93_ocj_AA300_12	HQ731908	Ocj	673
ripening_93_ocj_AA300_71a	HQ731907	Ocj	648
KS-dehydrin			
cold_47_ocj_AA300_12	HQ731848	Ocj	unknown
cold_47_os_PGS_44	HQ731846	Os	431
cold_47_ocj_AA300_71a_405plus	HQ731849	Ocj	unknown
cold_47_os_WAF_23d	HQ731845	Os	433
cold_47_ocj_AA300_71a_431bp	HQ731850	Ocj	431
cold_47_os_PGS_44g	HQ731847	Os	unknown
cold_47_oa_Aiken_sn_46e_433bp	HQ731839	Oa	433
cold_47_oa_Aiken_sn_65c	HQ731840	Oa	430
cold_47_oa_Aiken_sn_46e_430bp	HQ731842	Oa	430
cold_47_oa_Aiken_sn_46	HQ731841	Oa	430
cold_47_ocj_AA181_17a	HQ731844	Ocj	432
cold_47_om_Prevost_sn_101c	HQ731851	Om	unknown
cold_o_om_Prevost_sn_101c	HQ731843	Om	427
cold_59_M2B_oa_Prevost_sn_43b	HQ731837	Oa	383
cold_59_M2B_oa_Aiken_sn_46	HQ731863	Oa	156
cold_59_M2B_oa_Aiken_sn_65c	HQ731862	Oa	156
cold_59_M2B_om_Aiken_sn_20g	HQ731838	Om	unknown
cold_59_M2B_oa_Aiken_sn_46e	HQ731864	Oa	156
cold_59_M2B_om_Prevost_sn_101c	HQ731854	Om	unknown
cold_59_M1Y_om_Aiken_sn_20g_387bp	HQ731852	Om	387
cold_59_M1Y_om_Prevost_sn_101c	HQ731866	Om	382
cold_59_M1Y_ol_Sterling_s18f	HQ731860	Ol	394
cold_59_M1Y_os_WAF_88f_327bp	HQ731853	Os	327
cold_59_M1Y_os_PGS_44g	HQ731857	Os	384
cold_59_M1Y_os_PGS_44_384bp	HQ731856	Os	384
cold_59_M1Y_os_WAF_88f_374bp	HQ731858	Os	374
cold_59_M1Y_om_Aiken_sn_20g_385bp	HQ731855	Om	385
cold_59_M1Y_os_PGS_44_395bp	HQ731859	Os	395
cold_59_M1Y_ocj_AA300_12	HQ731861	Ocj	395
cold_59_M1Y_ocj_AA300_71a	HQ731865	Ocj	395
lhcaIII			
lhcaIII_oa_Aiken_sn_46e	HQ731875	Oa	99; 491
lhcaIII_oa_Aiken_sn_46	HQ731876	Oa	99; 491

lhcaIII_oa_Aiken_sn_65c	HQ731877	Oa	99; 491
lhcaIII_om_Aiken_sn_20g	HQ731878	Om	105; 471
lhcaIII_os_PGS_44g	HQ731879	Os	89; 429
lhcaIII_os_WAF_23d	HQ731880	Os	103; 200
lhcaIII_ocj_AA300_12	HQ731881	Ocj	95; 493
lhcaIII_ocj_AA300_71a	HQ731882	Ocj	95; 493

bp: base pairs

8.10 Appendix 10: Supplementary Table S4.4 List of sequence divergence[b] for confidently established pairs of alleles[a] and pairs of paralogs[a] from *Oxytropis* diploid genomes, and from pairs of genes with unclear relationships.

Sequence pairs with high sequence divergence[b] values from an *Oxytropis* plantlet may suggest they could be different gene copies, while low divergence[b] values may suggests they are different alleles of a single loci, and are indicated by an asterisk before the sequence name.

Gene family	Type of gene relationship[a]		Percent pairwise divergence[b]
PR-10	Alleles pairs		
	PR-10_13_os_PGS_44 HQ731803	PR-10_13_os_PGS_44g HQ731804	0
	PR-10_13_ocj_AA300_71a HQ731810	PR-10_13_ocj_AA300_12 HQ731811	0
	PR-10_f_om_Aiken_sn_40e HQ731830	PR-10_f_om_Aiken_sn_20g HQ731828	0
	PR-10_18_os_PGS_44g HQ731834	PR-10_18_os_WAF_88f HQ731833	0
	PR-10_18_oa_Aiken_sn_46 HQ731832	PR-10_18_oa_Prevost_sn_43b HQ731831	0
	PR-10_18_oa_Aiken_sn_65e HQ731836	PR-10_18_oa_Aiken_sn_46e HQ731829	0
	PR-10_61_oa_Aiken_sn_46e HQ731823	PR-10_61_oa_Aiken_sn_65e HQ731822	0
	PR-10_61_oa_Aiken_sn_65c HQ731821	PR-10_61_oa_Aiken_sn_65e HQ731822	0
	PR-10_61_oa_Aiken_sn_46e HQ731823	PR-10_61_oa_Aiken_sn_65c HQ731821	0
	PR-10_61_om_Aiken_sn_20g HQ731824	PR-10_61_om_Aiken_sn_40e HQ731825	1.3393
	PR-10_61_os_PGS_44g HQ731818	PR-10_61_os_PGS_44h HQ731817	0
	PR-10_61_os_PGS_44 HQ731816	PR-10_61_os_PGS_44h HQ731817	0
	PR-10_61_os_PGS_44g HQ731818	PR-10_61_os_PGS_44 HQ731816	0
	Paralog pairs		
	PR-10_13_oa_Aiken_sn_46_76bp HQ731802	PR-10_18_oa_Aiken_sn_46 HQ731832	8.4821
	PR-10_13_os_PGS_44g HQ731804	PR-10_18_os_PGS_44g HQ731834	7.1429
	PR-10_13_os_PGS_44g HQ731804	PR-10_61_os_PGS_44g HQ731818	8.4821
	PR-10_18_os_PGS_44g HQ731834	PR-10_61_os_PGS_44g HQ731818	5.3571
	PR-10_13_os_WAF_88f HQ731812	PR-10_18_os_WAF_88f HQ731833	7.5893
	PR-10_18_oa_Aiken_sn_46e HQ731829	PR-10_61_oa_Aiken_sn_46e HQ731823	5.8036
	PR-10_18_oa_Aiken_sn_65e HQ731836	PR-10_61_oa_Aiken_sn_65e HQ731822	5.8036
	Unclear gene relationship		
	PR-10_13_os_PGS_44 (HQ731803)	PR-10_f_os_PGS_44 (HQ731826)	6.6964
	PR-10_13_os_PGS_44 (HQ731803)	PR-10_61_os_PGS_44 (HQ731816)	8.4821
	PR-10_f_os_PGS_44 (HQ731826)	PR-10_61_os_PGS_44 (HQ731816)	4.9107
	PR-10_13_ocj_AA300_71a (HQ731810)	PR-10_61_ocj_AA300_71a_82bp (HQ731820)	3.5714
	PR-10_13_ocj_AA300_71a (HQ731810)	PR-10_61_ocj_AA300_71a_86bp (HQ731819)	3.5714

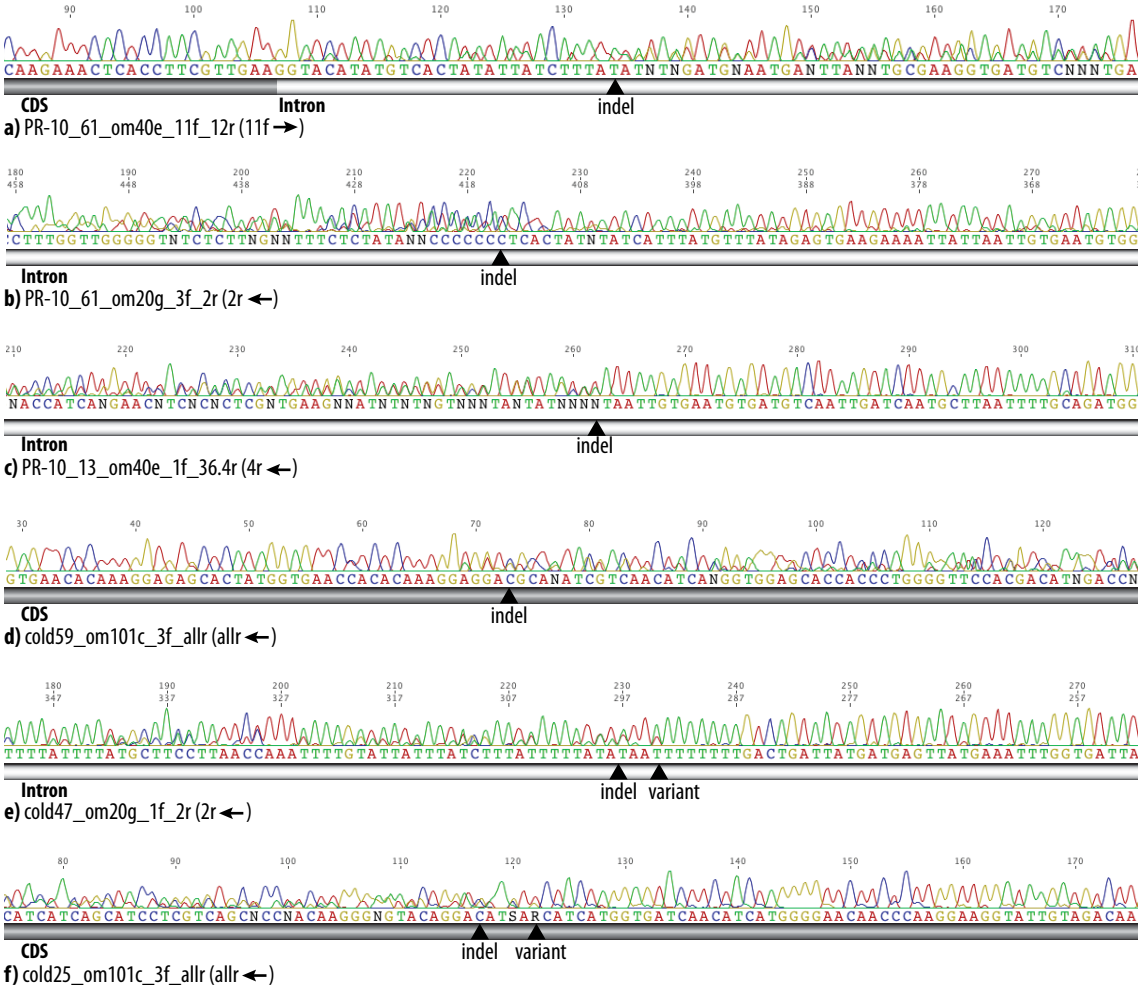
	* PR-10_61_ocj_AA300_71a_82bp (HQ731820)	PR-10_61_ocj_AA300_71a_86bp (HQ731819)	0
	PR-10_f_om_Aiken_sn_20g (HQ731828)	PR-10_61_om_Aiken_sn_20g (HQ731824)	4.0179
	PR-10_f_om_Aiken_sn_40e (HQ731830)	PR-10_61_om_Aiken_sn_40e (HQ731825)	3.5714
	* PR-10_13_oa_Aiken_sn_46_76bp (HQ731802)	PR-10_13_oa_Aiken_sn_46_99bp (HQ731808)	2.2321
	PR-10_13_oa_Aiken_sn_46_99bp (HQ731808)	PR-10_18_oa_Aiken_sn_46 (HQ731832)	6.6964
Ripening-related proteins	Allele pairs		
	ripening_93_oa_Aiken_sn_46e HQ731903	ripening_93_oa_Aiken_sn_65e HQ731904	0
	Paralogs pairs		
	ripening_93_oa_Aiken_sn_46e HQ731903	ripening_101_oa_Aiken_sn_46e HQ731901	19.3772
	ripening_32_os_PGS_44g HQ731896	ripening_101_os_PGS_44g HQ731902	19.7232
	ripening_p_oa_Aiken_sn_65c HQ731892	ripening_101_oa_Aiken_sn_46e HQ731901	23.1834
	ripening_p_oa_Aiken_sn_65c HQ731892	ripening_93_oa_Aiken_sn_65e HQ731904	24.2215
	ripening_p_oa_Aiken_sn_65c HQ731892	ripening_93_oa_Aiken_sn_46e HQ731903	24.2215
	Unclear gene relationship		
	ripening_p_om_Aiken_sn_20g (HQ731893)	ripening_32_om_Aiken_sn_20g (HQ731895)	26.6436
	ripening_93_ocj_AA300_71a (HQ731907)	ripening_p_ocj_AA300_71a (HQ731891)	25.9516
	ripening_101_om_Aiken_sn_40e (HQ731900)	ripening_93_om_Aiken_sn_40e (HQ731906)	19.0311
KS-dehydrins	Alleles pairs		
	cold_47_os_PGS_44 HQ731846	cold_47_os_WAF_23d HQ731845	0
	cold_47_os_PGS_44 HQ731846	cold_47_os_PGS_44g HQ731847	0
	cold_47_os_WAF_23d HQ731845	cold_47_os_PGS_44g HQ731847	0
	cold_47_oa_Aiken_sn_46e_433bp HQ731839	cold_47_oa_Aiken_sn_65c HQ731840	1.7391
	cold_47_oa_Aiken_sn_46e_433bp HQ731839	cold_47_oa_Aiken_sn_46 HQ731841	1.7391
	cold_47_oa_Aiken_sn_65c HQ731840	cold_47_oa_Aiken_sn_46e_430bp HQ731842	0
	cold_47_oa_Aiken_sn_65c HQ731840	cold_47_oa_Aiken_sn_46 HQ731841	0
	cold_59_M2B_oa_Aiken_sn_46 HQ731863	cold_59_M2B_oa_Aiken_sn_65c HQ731862	0
	cold_59_M2B_oa_Aiken_sn_46 HQ731863	cold_59_M2B_oa_Aiken_sn_46e HQ731864	0
	cold_59_M2B_oa_Aiken_sn_65c HQ731862	cold_59_M2B_oa_Aiken_sn_46e HQ731864	0
	cold_59_M1Y_os_WAF_88f_327bp HQ731853	cold_59_M1Y_os_PGS_44g HQ731857	0
	cold_59_M1Y_os_WAF_88f_327bp HQ731853	cold_59_M1Y_os_PGS_44_384bp HQ731856	0
	cold_59_M1Y_os_WAF_88f_327bp HQ731853	cold_59_M1Y_os_PGS_44_395bp HQ731859	0
	cold_59_M1Y_os_PGS_44g HQ731857	cold_59_M1Y_os_PGS_44_384bp HQ731856	0
	cold_59_M1Y_os_PGS_44g HQ731857	cold_59_M1Y_os_WAF_88f_374bp HQ731858	0
	cold_59_M1Y_os_PGS_44g HQ731857	cold_59_M1Y_os_PGS_44_395bp HQ731859	0

cold_59_M1Y_os_PGS_44_384bp HQ731856	cold_59_M1Y_os_WAF_88f_374bp HQ731858	0
cold_59_M1Y_os_WAF_88f_374bp HQ731858	cold_59_M1Y_os_PGS_44_395bp HQ731859	0
Paralog pairs		
cold_47_oa_Aiken_sn_46e_433bp HQ731839	cold_59_M2B_oa_Aiken_sn_46e HQ731864	8.6957
cold_47_oa_Aiken_sn_65c HQ731840	cold_59_M2B_oa_Aiken_sn_65c HQ731862	8.6957
cold_47_oa_Aiken_sn_46e_430bp HQ731842	cold_59_M2B_oa_Aiken_sn_46e HQ731864	8.6957
cold_47_oa_Aiken_sn_46 HQ731841	cold_59_M2B_oa_Aiken_sn_46 HQ731863	8.6957
cold_47_os_PGS_44 HQ731846	cold_59_M1Y_os_PGS_44_384bp HQ731856	5.2174
cold_47_os_PGS_44g HQ731847	cold_59_M1Y_os_PGS_44g HQ731857	5.2174
cold_47_os_WAF_23d HQ731845	cold_59_M1Y_os_WAF_88f_327bp HQ731853	5.2174
cold_47_os_WAF_23d HQ731845	cold_59_M1Y_os_WAF_88f_374bp HQ731858	5.2174
cold_59_M1Y_os_PGS_44_395bp (HQ731859)	cold_47_os_PGS_44 (HQ731846)	5.2174
Unclear gene relationship		
* cold_47_oa_Aiken_sn_46e_433bp HQ731839	cold_47_oa_Aiken_sn_46e_430bp HQ731842	1.7391
* cold_59_M1Y_os_WAF_88f_327bp HQ731853	cold_59_M1Y_os_WAF_88f_374bp HQ731858	0
* cold_59_M1Y_os_PGS_44_384bp HQ731856	cold_59_M1Y_os_PGS_44_395bp HQ731859	0
cold_47_ocj_AA300_12 (HQ731848)	cold_59_M1Y_ocj_AA300_12 (HQ731861)	5.2174
cold_47_ocj_AA300_71a_431bp (HQ731850)	cold_59_M1Y_ocj_AA300_71a (HQ731865)	6.087
* cold_47_ocj_AA300_71a_431bp (HQ731850)	cold_47_ocj_AA300_71a_405plus (HQ731849)	0
cold_59_M1Y_ocj_AA300_71a (HQ731865)	cold_47_ocj_AA300_71a_405plus (HQ731849)	6.087
cold_59_M2B_om_Aiken_sn_20g (HQ731838)	cold_59_M1Y_om_Aiken_sn_20g_385bp (HQ731855)	4.7826
cold_59_M2B_om_Aiken_sn_20g (HQ731838)	cold_59_M1Y_om_Aiken_sn_20g_387bp (HQ731852)	4.7826
cold_59_M1Y_om_Aiken_sn_20g_385bp (HQ731855)	cold_59_M1Y_om_Aiken_sn_20g_387bp (HQ731852)	3.4783
cold_59_M1Y_os_PGS_44_395bp (HQ731859)	cold_47_os_PGS_44 (HQ731846)	5.2174
cold_59_M1Y_os_PGS_44g (HQ731857)	cold_47_os_PGS_44g (HQ731847)	5.2174
cold_59_M1Y_om_Prevost_sn_101c (HQ731866)	cold_59_M2B_om_Prevost_sn_101c (HQ731854)	1.3043
cold_59_M2B_om_Prevost_sn_101c (HQ731854)	cold_o_om_Prevost_sn_101c (HQ731843)	4.3478
cold_59_M1Y_om_Prevost_sn_101c (HQ731866)	cold_o_om_Prevost_sn_101c (HQ731843)	5.6522
cold_59_M1Y_om_Prevost_sn_101c (HQ731866)	cold_47_om_Prevost_sn_101c (HQ731851)	6.5217
cold_59_M2B_om_Prevost_sn_101c (HQ731854)	cold_47_om_Prevost_sn_101c (HQ731851)	5.2174
* cold_o_om_Prevost_sn_101c (HQ731843)	cold_47_om_Prevost_sn_101c (HQ731851)	0.8696

[a] The relationship of pairs of alleles and pairs of paralogs is confidently established for diploid genomes (*O. arctobia* and *O. splendens*) from the phylogenetic trees in Figure 4.2. Confident alleles pairs are from two different plantlets of a species, and are located in the same terminal branch of the phylogenetic tree. Confident paralogs pairs are from the same plantlet, and are located in two different estimated clades of the phylogenetic trees in Figure 4.2.

[b] The percentage divergence was based from a complete deletion, computed in MEGA (Kumar et al. 2008). PR-10 distances were computed from 224 positions, cold dehydrin from 230 positions, and ripening-related proteins from 289 positions.

* Suspected alleles



8.11 Appendix 11: Supplementary Figure S4.1 Chromatograms showing double peaks for some PCR amplified fragments from gDNA of the arctic dodecaploid *Oxytropis maydelliana.*

Chromatograms have clear single peaks until an indel, after which double peaks occur at many or most positions. Fragments were amplified using primer pairs for the PR-10 and the KS-dehydrins gene families: a) PR-10_61_11F with PR-

10_61_12R primers on plantlet Om40e gDNA b) PR-10_61_3F with PR-10_61_2F primers on plantlet Om20g gDNA c) PR-10_13_1F with PR-10_36_4R primers on plantlet om40e gDNA d) cold59_3F with cold_allR primers on plantlet Om101c gDNA e) cold47_1F with cold47_2R primers on plantlet Om20g gDNA f) cold25_3F with cold_allR primers on plantlet Om101c gDNA.

8.12 Appendix 12: Supplementary Table S4.5 Characteristics of primers and probes used for the estimation of copy number by qPCR experiment for PR-10 and KS-dehydrin gene families in *Oxytropis* species.

Target gene	Probe; sequence (5' to 3')	Forward primer; sequence (5' to 3')	Reverse primer; sequence (5' to 3')	Amplicon length (bp)	Used for final assay
KS-dehydrin	dehydrin_BHQprobe_4_All 5' 6-FAM-GTAGACAAGATCAAAGACAAGRTCCATGGTGA-BHQ1 3'	dehydrin_All_19F 5' CCATGGTATCATTGGTGGTGAACAC 3'	dehydrin_All_9R 5' TCCCWCTCCAYTTTCACCATCA 3'	247, 271, 301	No
KS-dehydrin	dehydrin_BHQprobe_2_All 5' 6-FAM-AAKAAGAAACATGAACATGGTCATGA-BHQ1 3'	dehydrin_All_5F; 5' GGTGATGGTGAAARTGGAGWGGGA 3'	dehydrin_All_8R; 5' AGAAGAAGCAAGATCTAATCACTSTC 3'	92, 104, 107, 110, 122	Yes
Pathogenesis related class 10	PR10_BHQprobe_3_All; 5' 6-FAM-GCTTTACAAAGCTCTTGTTAAGGATGC-BHQ1 3'	PR10_All_1F; 5' TYTYCACATTTGAGCAAGAAAC 3'	PR10_All_2R; 5' GGRCCACCATTTCCYTCAAC 3'	140	No
Pathogenesis related class 10	PR10_BHQprobe_5_All; 5' 6-FAM-ATTGTTGGTGGAGTTGGGTTACCAG-BHQ1 3'	PR10_All_7F; 5' TGTGTTGCACAAAGTAGARGCA 3'	PR10_All_8R; 5' CCRTCRGGGCCTGCAAWCAATT 3'	132	Yes
snrnap (small nuclear ribonucleoprotein-associated protein B)	snrnap_BHQprobe_4_All; 5' 6-FAM-CAAAGATGGGAGGTTCTTCTTGGGTAGCT-BHQ1 3'	snrnap_All_5F; 5' GCTGTTTCGTCGAATGTTGGT 3'	snrnap_All_6R; 5' CCATTGGAGAAGGCGATGAT 3'	143	Yes

Tlp15 (thylakoid lumenal 15.0 kDa protein	thyllum_BHQprobe_3_All 5’ 6-FAM-AGCTGGCGTGAACAAACCAGA-BHQ1 3’	thyllum_All_6F; 5’ TTATTATTCGGAGGAACTGAATTGG 3’	thyllum_All_8R; 5’ GACATCAATGACTGCGGTAAATTCT 3’	96	No

8.13 Appendix 13: Supplementary Table S4.6 Characteristics of the standard curves for qPCR reactions performed to estimate copy number of the target gene families PR-10 and KS-dehydrin, relatively to the reference gene snrnp, on two diploid (*O. arctobia* and *O. splendens*), and two polyploids (*O. maydelliana* and *O. campestris johannensis*) *Oxytropis* species.

Plantlet used as template DNA	Gene	Efficiency	Y-intercept	R^2
O. arctobia Oa57a	snrnp	110.004	27.0623	0.882
O. arctobia Oa57a	PR-10	97.378	25.9575	0.918
O. arctobia Oa57a	KS-dehydrin	104.093	27.1114	0.978
O. arctobia Oa65b	Snrnp	110.64	26.648	0.831
O. arctobia Oa65b	PR-10	88.855	26.1836	0.937
O. arctobia Oa65b	KS-dehydrin	109.509	27.2323	0.956
O. splendens Os54a	Snrnp	115.295	27.3609	0.966
O. splendens Os54a	PR-10	91.668	25.8238	0.986
O. splendens Os54a	KS-dehydrin	106.383	27.428	0.986
O. splendens Os99a	Snrnp	105.164	26.813	0.938
O. splendens Os99a	PR-10	103.465	25.9783	0.991
O. splendens Os99a	KS-dehydrin	104.428	27.2525	0.987
O. maydelliana Om101b	Snrnp	91.64	26.624	0.91
O. maydelliana Om101b	PR-10	98.52	24.912	0.99
O. maydelliana Om101b	KS-dehydrin	93.07	27.018	0.95
O. maydelliana Om47b	Snrnp	94.58	27.769	0.94
O. maydelliana Om47b	PR-10	92.46	24.806	0.98
O. maydelliana Om47b	KS-dehydrin	95.45	27.125	0.97

O. campestris johannensis Ocj87a	Snrnp	91.60	26.319	0.94
O. campestris johannensis Ocj87a	PR-10	93.32	25.213	0.99
O. campestris johannensis Ocj87a	KS-dehydrin	91.18	27.239	0.91
O. campestris johannensis Ocj98a	Snrnp	99.91	27.364	0.88
O. campestris johannensis Ocj98a	PR-10	96.03	24.838	0.98
O. campestris johannensis Ocj98a	KS-dehydrin	93.91	27.414	0.95

The standard curves were built with the Viia7 analysis software from the Ct values for each of the 6-fold serial dilutions of ½ total DNA.

8.14 Appendix 14: Supplementary Table S5.1 Seed sources, with localities and collectors, used to grow plantlets to provide DNA material of *Oxytropis* species for the sequences evolutionary relationships analyses.

Oxytropis species	Locality	Collector	Herbarium specimen
From the Arctic region			
O. arctobia Bunge	Baffin Island (Nunavut, Canada)	Dr. Susan Aiken, Canadian Museum of Nature, Ottawa	MTMG 136647 (Aiken SA7)
O. arctobia	Melville Peninsula (Nunavut, Canada)	Dr. Danielle Prévost, Agriculture and Agri-Food Canada, Ste-Foy (Prevost et al. 1987)	-
O. deflexa (Pall.) DC. subsp. *foliolosa* (Hook.) Cody	Baffin Island, Nunavut, Canada	Annie Archambault	AA298; AA277 [a]
O. maydelliana	Melville Peninsula Nunavut, Canada	Dr. Danielle Prévost, Agriculture and Agri-Food Canada, Ste-Foy (Prevost et al. 1987)	-
O. maydelliana	Baffin Island (Nunavut, Canada)	Annie Archambault	MTMG 136650 (AA214)
O. maydelliana	Seward Peninsula (Alaska), USA	Carolyn Parker, University of Alaska Fairbanks Museum of the North	ALA V97877
O. podocarpa A. Gray	Baffin Island, Nunavut, Canada	Annie Archambault	MTMG 136648 (AA207); MTMG 136649 (AA222)
From the temperate regions			
O. campestris (L.) DC subsp. *johannensis* (Fernald) Blondeau & Gervais	Ile d'Orléans, Quebec, Canada	Annie Archambault	MTMG 136652 AA181
O. campestris (L.) DC	Labrador, Newfoundland (Canada)	Dr. Laurie Consaul, Canadian Museum of	LC32353

		Nature, Ottawa	
O. splendens Douglas	Southern Alberta, Canada	Plant Gene Resources of Canada (Ottawa) (CN105143)	-
O. splendens	Longview, Southern Alberta, Canada	Wild About Flowers, native plant nursery	-
O. splendens	Cochin, Southern Saskatchewan, Canada	Prairie Garden Seeds, native plant nursery	-
From nearly desert region			-
O. lambertii Pursh	Ocate, New Mexico, USA	Dr. Tracy Sterling New Mexico State University, Las Cruces (Kulshreshtha et al. 2004)	-

ALA: University of Alaska Museum Herbarium

MTMG: McGill University, Macdonald Campus

[a] There is no voucher for these two plantlets, but there is a voucher for sample AA211, an *Oxytropis deflexa* subsp. *foliolosa* collected by Annie Archambault from the same locality, on the same season; its accession number is MTMG 136651.

8.15 Appendix 15: Supplementary Table S5.2 Genbank accession number for the *Oxytropis* sequences of the nuclear ribosomal internal transcribed spacer (ITS) included in the sequences evolutionary relationships analyses.

Oxytropis species	Geographical origin of specimen	GenBank accession number	Subgenus	Section	Source
O. arctobia	North America; Tundra (Canada, Dorset Island)	HQ176477	*Oxytropis*	*Arctobia*	This study
O. arctobia	North America; Tundra (Canada, Melville Peninsula)	HQ176487	*Oxytropis*	*Arctobia*	This study
O. arctobia	North America; Tundra (Canada, Baffin Island)	HQ176476	*Oxytropis*	*Arctobia*	This study
O. nigrescens	North America; Tundra (USA, Alaska)	AF366348 and AF366349	*Oxytropis*	*Arctobia*	(Jorgensen et al. 2003)
O. podocarpa	North America; Tundra (Canada, Baffin Island)	HQ176480	*Oxytropis*	*Arctobia*	This study
O. podocarpa	North America; Tundra (Canada, Baffin Island)	HQ176483	*Oxytropis*	*Arctobia*	This study
O. borealis var. *viscida*	North America; Desert (USA, Utah)	AF121758	*Oxytropis*	*Gloeocephala*	(Wojciechowski et al. 1999)
O. arctica var. *barnebyana*	North America; Tundra (USA, Alaska)	AF366304 and AF366305	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. arctica var. *barnebyana*	North America; Tundra (USA, Alaska)	AF366306 and AF366307	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. arctica var. *barnebyana*	North America; Tundra (USA, Alaska)	AF366302 and AF366303	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. arctica var. *barnebyana*	North America; Tundra (USA, Alaska)	AF366298 and AF366300	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. arctica var. *barnebyana*	North America; Tundra (USA, Alaska)	AF366308 and AF366309	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. arctica var. *barnebyana*	North America; Tundra (USA, Alaska)	AF366310 and AF366311	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)

O. arctica var. *koyokukensis*	North America; Tundra (USA, Alaska)	AF366332 and AF366333	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. arctica var. *koyokukensis*	North America; Tundra (USA, Alaska)	AF366330 and AF366331	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. arctica var. *barnebyana*	North America; Tundra (USA, Alaska)	AF366312 and AF366313	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. arctica var. *koyukukensis*	North America; Tundra (USA, Alaska)	AF366334 and AF366335	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. arctica var. *arctica*	North America; Tundra (USA, Alaska)	AF366326 and AF366327	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. arctica var. *arctica*	North America; Tundra (USA, Alaska)	AF366324 and AF366325	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. besseyi var. *ventosa*	North America; Desert (USA, Utah)	AF121756	*Oxytropis*	*Orobia*	(Wojciechowski et al. 1999)
O. campestris	North America; Temperate boreal forest (Canada, Labrador)	HQ176475	*Oxytropis*	*Orobia*	This study
O. campestris subsp. *gracilis*	North America; Tundra (USA, Alaska)	AF366342 and AF366343	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. campestris subsp. *gracilis*	North America; Tundra (USA, Alaska)	AF366340 and AF366341	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. campestris subsp. *gracilis*	North America; Tundra (USA, Alaska)	AF366338 and AF366339	*Oxytropis*	*Orobia*	(Jorgensen et al. 2003)
O. campestris var. *johannensis*	North America; Temperate deciduous forest (USA, Maine)	L10802	*Oxytropis*	*Orobia*	(Wojciechowski et al. 1993)
O. campestris subsp. *johannensis*	North America; Temperate deciduous forest (Canada, Québec)	HQ176478	*Oxytropis*	*Orobia*	This study
O. lambertii	North America; Desert (USA, New Mexico)	HQ176472	*Oxytropis*	*Orobia*	This study
O. lambertii	North America; Desert (USA, Utah)	L10807	*Oxytropis*	*Orobia*	(Wojciechowski et al. 1993)
O. lambertii	North America; Desert (USA, Arizona)	AF121753	*Oxytropis*	*Orobia*	(Wojciechowski et al. 1999)
O. maydelliana	North America; Tundra (Canada, Melville peninsula)	HQ176485	*Oxytropis*	*Orobia*	This study
O. maydelliana	North America; Tundra (USA, Alaska)	HQ176486	*Oxytropis*	*Orobia*	This study
O. maydelliana	North America; Tundra (Canada, Baffin Island)	HQ176484	*Oxytropis*	*Orobia*	This study
O. sericea	North America; Desert (USA, Utah)	AF121757	*Oxytropis*	*Orobia*	(Wojciechowski et al. 1999)
O. ochrantha	Central Asia; Temperate deciduous forest (China, Huitengliang steppes)	GQ422819	*Oxytropis*	*Verticillares*	(Gao et al. 2009)

O. ochrantha	Central Asia; Temperate deciduous forest (China, Zhuozi County)	GQ422820	*Oxytropis*	*Verticillares*	(Gao et al. 2009)
O. racemosa	Central Asia; Temperate deciduous forest (China, Etuoke Banner)	GQ422811	*Oxytropis*	*Verticillares*	(Gao et al. 2009)
O. racemosa	Central Asia; Temperate deciduous forest (China, Wushen Banner)	GQ422812	*Oxytropis*	*Verticillares*	(Gao et al. 2009)
O. racemosa	Central Asia; Temperate deciduous forest (China, Yijinhuolo Banner)	GQ422813	*Oxytropis*	*Verticillares*	(Gao et al. 2009)
O. racemosa	Central Asia; Temperate deciduous forest (China, Helingeer County)	GQ422814	*Oxytropis*	*Verticillares*	(Gao et al. 2009)
O. racemosa	Central Asia; Temperate deciduous forest (China, Hang Jin Qi)	HQ199319	*Oxytropis*	*Verticillares*	Unpublished
O. racemosa	Central Asia; Temperate deciduous forest (China, Qian Qi)	HQ199320	*Oxytropis*	*Verticillares*	Unpublished
O. splendens	North America; Temperate grasslands (USA, Colorado)	AF121761	*Oxytropis*	*Verticillares*	(Wojciechowski et al. 1999)
O. splendens	North America; Temperate grasslands (Canada, Alberta)	HQ176474	*Oxytropis*	*Verticillares*	This study
O. splendens	North America; Temperate grasslands (Canada, Saskatchewan)	HQ176473	*Oxytropis*	*Verticillares*	This study
O. splendens	North America; Temperate grasslands (Canada, Saskatchewan)	HQ176479	*Oxytropis*	*Verticillares*	This study
O. verticillaris	Central Asia; Temperate deciduous forest (China, Wuchuan County)	GQ422815	*Oxytropis*	*Verticillares*	(Gao et al. 2009)
O. verticillaris	Central Asia; Temperate deciduous forest (China, Wuchuan County)	GQ422816	*Oxytropis*	*Verticillares*	(Gao et al. 2009)
O. verticillaris	Central Asia; Temperate deciduous forest (China, Wuchuan County)	GQ422817	*Oxytropis*	*Verticillares*	(Gao et al. 2009)
O. verticillaris	Central Asia; Temperate deciduous forest (China, Wuchuan County)	GQ422818	*Oxytropis*	*Verticillares*	(Gao et al. 2009)
O. bicolor	Central Asia; Temperate deciduous forest (China)	HQ199317	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839005	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839010	*Oxytropis*	*Verticillares*	Unpublished

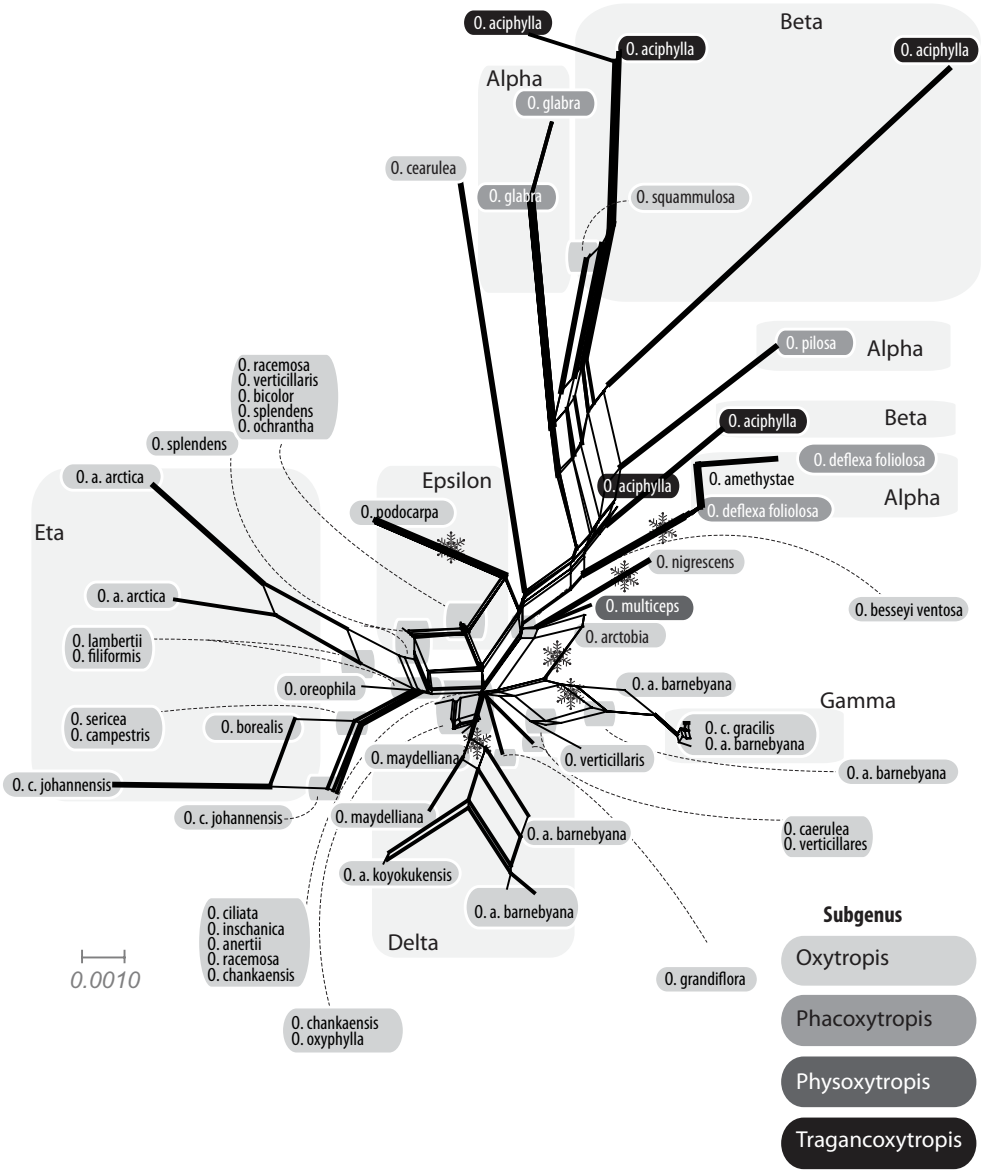
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839001	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839004	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839012	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839008	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839006	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839018	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839007	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839009	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839003	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839013	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839017	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839015	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839014	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839002	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839016	*Oxytropis*	*Verticillares*	Unpublished
O. chankaensis	Central Asia; Temperate deciduous forest (Russia, Khanka Lake, Primorye)	FR839011	*Oxytropis*	*Verticillares*	Unpublished
O. oxyphylla	Central Asia; Temperate deciduous forest (Russia,Zabaykalsky Krai, Chita)	FR839000	*Oxytropis*	*Verticillares*	Unpublished
O. grandiflora	Central Asia; Temperate deciduous forest (China)	HQ199315	*Oxytropis*	*Oxytropis*	Unpublished

O. anertii	Central Asia; temperate deciduous forest (China, Tonghua, Jilin)	EF685971	*Oxytropis*	*Xerobia*	(Guo et al. 2010)
O. ciliata	Central Asia; temperate deciduous forest (China, Huang Hua Gacha)	HQ199325	*Oxytropis*	*Xerobia*	Unpublished
O. ciliata	Central Asia; temperate deciduous forest (China, Wah Xing Yuan)	HQ199323	*Oxytropis*	*Xerobia*	Unpublished
O. ciliata	Central Asia; temperate deciduous forest (China, Wu Sheng Guandi)	HQ199324	*Oxytropis*	*Xerobia*	Unpublished
O. inschanica	Central Asia; temperate deciduous forest (China, Shi Er Deng)	HQ199322	*Oxytropis*	*Xerobia*	Unpublished
O. caerulea	Central Asia; temperate deciduous forest (China)	HQ199316	*Oxytropis*	*Eumorpha*	Unpublished
O. caerulea	Central Asia; temperate deciduous forest (China)	GU217599	*Oxytropis*	*Eumorpha*	Unpublished
O. filiformis	Central Asia; temperate deciduous forest (China, Wu Sheng Guandi)	HQ199321	*Oxytropis*	*Eumorpha*	Unpublished
O. squammulosa	Central Asia; temperate deciduous forest (China, Wu Sheng Guandi)	HQ199318	*Oxytropis*	*Leucopodia*	Unpublished
O. oreophila	North America; Desert (USA, Utah)	AF121755	*Phacoxytropis*	*Janthina*	(Wojciechowski et al. 1999)
O. amethystea	Europe; temperate deciduous forest	GQ246045	Unknown	Unknown	Unpublished
O. deflexa var. sericea	North America; Temperate grasslands (USA, Colorado)	L10804	*Phacoxytropis*	*Mesogaea*	(Wojciechowski et al. 1993)
O. deflexa subsp. foliolosa	North America; Tundra (Canada, Baffin Island)	HQ176481	*Phacoxytropis*	*Mesogaea*	This study
O. deflexa subsp. foliolosa	North America; Tundra (Canada, Baffin Island)	HQ176482	*Phacoxytropis*	*Mesogaea*	This study
O. pilosa	Northern Asia; Temperate deciduous forest (Kazakh territory)	AF121759	*Phacoxytropis*	*Mesogaea*	(Wojciechowski et al. 1999)
O. glabra	Central Asia; Temperate deciduous forest (China, Inner Mongolia, Alxazuo Banner)	GQ265958	*Phacoxytropis*	*Mesogaea*	(Gao et al. 2009)
O. glabra	Central Asia; Temperate deciduous forest (China, Inner Mongolia, Linhe District)	GQ265959	*Phacoxytropis*	*Mesogaea*	(Gao et al. 2009)
O. glabra	Central Asia; Temperate deciduous forest (China, Inner Mongolia Wushen Banner)	GQ265960	*Phacoxytropis*	*Mesogaea*	(Gao et al. 2009)
O. glabra	Central Asia; Temperate deciduous forest (China, Inner Mongolia Yijinhuolo Banner)	GQ265961	*Phacoxytropis*	*Mesogaea*	(Gao et al. 2009)

O. glabra	Central Asia; temperate deciduous forest (China, Inner Mongolia, Etuoqian Banner)	GQ265962	*Phacoxytropis*	*Mesogaea*	(Gao et al. 2009)
O. glabra	Central Asia; Temperate deciduous forest (China, Yijinhuolo Banner)	GQ422805	*Phacoxytropis*	*Mesogaea*	(Gao et al. 2009)
O. multiceps	North America; Temperate grasslands (USA, Colorado)	AF121760	*Physoxytropis*	*Physoxytropis*	(Wojciechowski et al. 1999)
O. aciphylla	Central Asia; Temperate deciduous forest (China,Yijinhuolo Banner)	GQ422807	*Tragacanthoxytropis*	*Lycotriche*	(Gao et al. 2009)
O. aciphylla	Central Asia; Temperate deciduous forest (China, Zhungeer Banner)	GQ422810	*Tragacanthoxytropis*	*Lycotriche*	(Gao et al. 2009)
O. aciphylla	Central Asia; Temperate deciduous forest (China, Wulatezhong Banner)	GQ422806	*Tragacanthoxytropis*	*Lycotriche*	(Gao et al. 2009)
O. aciphylla	Central Asia; Temperate deciduous forest (China, Etuoke Banner)	GQ422808	*Tragacanthoxytropis*	*Lycotriche*	(Gao et al. 2009)
O. aciphylla	Central Asia; Temperate deciduous forest (China, Alxazuo Banner)	GQ422809	*Tragacanthoxytropis*	*Lycotriche*	(Gao et al. 2009)

Note: The section *Verticillares* (Malyshev 2009), corresponds to section *Baicalia* (Langran et al. 2010), and the *O. chankaensis* is synonymous to *O. oxyphylla* (Langran et al. 2010).

8.16 Appendix 16: Supplementary Figure S5.1 Relationships among 30 *Oxytropis* species (97 sequences) calculated with a split network on the nuclear ribosomal internal transcribed spacer (ITS) sequences.



Underlying light grey boxes indicate informal groups Alpha to Eta. The two or three letter code refers to the taxonomic section of the species Ar: *Arctobia*, Eu: *Eumorpha*, Gl: *Gloeocephala*, Ja: *Janthina*, Le: *Leucopodia*, Ly: *Lycotriche*, Me: *Mesogaea*, Or: *Orobia*; Ox: *Oxytropis*, Phy: *Physoxytropis*, Ve: *Verticillares*, Xe: *Xerobia*. Thickness of an edge is proportional to its bootstrap support.